UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS
Departamento de Genética

TESIS DOCTORAL

Identificación genómica de loci y rutas génicas que afectan al crecimiento y deposición grasa en porcino

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Dafne Pérez Montarelo

Directora

Ana I. Fernández Ávila

Madrid, 2014

UNIVERSIDAD COMPLUTENSE DE MADRID

Identificación genómica de loci y rutas génicas que afectan al crecimiento y deposición grasa en porcino

Dafne Pérez Montarelo

TESIS DOCTORAL

Octubre 2013

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE BIOLOGÍA

Departamento de Genética

Identificación genómica de loci y rutas génicas que afectan al crecimiento y deposición grasa en porcino

Genomic identification of loci and genetic pathways affecting growth and fat deposition in swine

Tesis Doctoral

Dafne Pérez Montarelo

VºBº Directora VºBº Tutora La Doctoranda

Ana I. Fernández Ávila Julia Rueda Muñoz Dafne Pérez Montarelo

Madrid, Octubre de 2013

ÍNDICE

RESUMEN

La carne roja de mayor consumo mundial es la carne de cerdo, por tanto la producción porcina constituye el subsector pecuario de mayor crecimiento, donde España es el segundo mayor productor en Europa después de Alemania. En este contexto, el crecimiento y la deposición grasa son algunos de los caracteres económicos más relevantes en la producción porcina. Además, el cerdo es utilizado como especie modelo en el estudio de la obesidad y patologías humanas relacionadas con ella, dada su mayor semejanza en anatomía, genética y fisiopatología. Por todo ello, el estudio de estos caracteres suscita un gran interés desde el punto de vista tanto biomédico como productivo.

Diversos factores contribuyen a la variación fenotípica del crecimiento y la deposición grasa. La composición de la dieta, la edad y el sexo tienen un gran efecto sobre estos. Asimismo, factores medioambientales como los sistemas de producción, las horas de luz y la temperatura influyen en los requerimientos individuales afectando al contenido graso de los animales. Por otro lado, las diferencias existentes en el crecimiento y la deposición grasa de las distintas razas porcinas indican la importancia de los factores genéticos en la regulación de estos procesos. La heredabilidad de estos caracteres en porcino se ha estimado en aproximadamente 0.45 para la deposición grasa y 0.25 para caracteres de crecimiento.

El objetivo general de esta tesis ha sido profundizar en el conocimiento de la base genética de la regulación del crecimiento y la deposición grasa en porcino, identificando genes, rutas génicas y mutaciones que afecten a estos caracteres. Este objetivo general ha sido abordado desde distintas perspectivas utilizando tanto metodologías de análisis masivos como metodologías más tradicionales sobre el material animal de un cruce experimental de cerdo Ibérico x Landrace.

El cruce experimental Ibérico x Landrace fue desarrollado en 1996 con el objetivo de estudiar la base genética del crecimiento, deposición grasa, composición corporal y calidad de la carne en porcino. En principio se obtuvo una generación F2 y posteriormente se obtuvo nuevo material animal en las generaciones F3 y dos retrocruces procedentes del cruce de machos F2 x Landrace y de machos F1 x Landrace. Entre los resultados más relevantes obtenidos en este material y relacionados con el crecimiento y la deposición grasa cabe destacar la

detección de QTLs en los cromosomas SSC2, SSC4 y SSC6. El QTL localizado en el SSC6 ha sido estudiado en mayor profundidad identificando al gen *LEPR* como candidato biológico y posicional a portar la mutación causal de los efectos de este QTL. De hecho, se identificó una mutación puntual, *LEPRc*.1987C>T, fuertemente asociado a caracteres de deposición grasa y crecimiento, y se determinó que este SNP afecta al nivel de expresión del gen *LEPR* en hipotálamo, sin embargo no existen evidencias del efecto causal de dicha mutación. A partir de estos resultados previos relacionados con el crecimiento y la deposición grasa se ha desarrollado la presente tesis doctoral.

Se han utilizado dos aproximaciones complementarias de análisis masivo del genoma sobre el material animal del cruce experimental Ibérico x Landrace para conseguir abordar el estudio de estos caracteres. Una de ellas consistió en análisis de detección de regiones QTL mediante análisis de ligamiento explotando la información del chip de genotipado masivo de SNPs realizado sobre la generación F3 y los dos retrocruces de la población IBMAP. Los análisis se realizaron utilizando la información conjunta de las generaciones F3 y los dos retrocruces o separándola de acuerdo a la generación machos utilizados para su obtención, poniendo de manifiesto las diferencias entre los genotipos para los QTL esperados y el muestreo aleatorio de los alelos de los QTLs en los individuos F1 y F2. Los resultados de estos análisis ha permitido corroborar e identificar nuevas regiones del genoma asociadas a los caracteres de interés, así como proponer potentes genes candidatos dentro de las mismas. Concretamente, se han confirmado los QTLs localizados en SSC4 y SSC6 para deposición grasa y con intervalos de confianza que se han visto reducidos de 12 y 18 cM a solo 7 y 2 cM, respectivamente. Adicionalmente, se han identificado nueve nuevos QTLs en los cromosomas SSC1, SSC2, SSC4, SSC5, SSC9, SSC11 SSC13, SSC14 y SSC17 asociados a distintos caracteres de crecimiento y deposición grasa, pudiéndose identificar en la mayoría de ellos potentes genes candidatos posicionales y funcionales, como los genes *PTPRD, VCAM1*, *AGL*, *ADIPOR2, VDR, MYOG, FASLG, NCK1, CYP2E1* e *ID1*.

Por otro lado, se llevó a cabo un análisis del transciptoma hipotalámico porcino, en animales con fenotipos divergentes para crecimiento y deposición grasa de uno de los retrocruces de la población IBMAP mediante la tecnología RNA-Seq. El

hipotálamo es el principal tejido regulador de la ingesta de alimentos, el balance energético y el peso corporal. En este trabajo se ha mostrado la elevada complejidad de este tejido en comparación con otros como el hígado y las gónadas y su importante implicación en los caracteres de interés. Además, mediante un análisis de expresión diferencial, se han podido detectar más de 200 transcritos conocidos y 50 nuevas isoformas diferencialmente expresadas entre los dos grupos de individuos divergentes para crecimiento y deposición grasa. La anotación biológica y funcional de éstos ha permitido identificar genes, tránscritos y potenciales reguladores transcripcionales involucrados en rutas relacionadas con el control del crecimiento y la deposición grasa. Así mismo, se han identificado potentes genes candidatos que se localizan dentro de las regiones QTL identificadas en el estudio anterior. De la lista de genes e isoformas diferencialmente expresadas identificados mediante esta estrategia destacan los genes *EGR1* e *IRF1*, por ser tanto candidatos posicionales, localizados en el QTL SSC2, como funcionales. El gen *IRF1* codifica un factor de transcripción involucrado en la regulación de la hormona de crecimiento. El *EGR1* codifica así mismo un factor de trancripción que juega un importante papel en la regulación de la ingesta de alimentos y de la duración y cantidad de la ingesta. Junto con ellos, aunque no aparezca diferencialmente expresado en este estudio, también destaca el gen *NR3C1*, ya que se trata de un factor de transcripción que regula la expresión de cinco genes identificados como diferencialmente expresados en el análisis.

Además, se han utilizado estrategias de análisis más tradicionales para el estudio de los genes candidatos: *leptina (LEP*) y *receptor de la leptina (LEPR*). Ambos genes son especialmente relevantes en este estudio dada su importante función sobre la regulación del crecimiento y deposición grasa y los resultados previos del polimorfismo *LEPR*c.1987C>T en porcino. La leptina es una hormona periférica reguladora de la saciedad producida principalmente por los adipocitos en función de la cantidad de materia grasa. Es liberada al torrente sanguíneo a través del cual llega a su receptor específico (LEPR) localizado principalmente en el hipotálamo produciendo su activación y desencadenando una cascada de señalización que produce sensación de saciedad. Debido a la falta de anotación del genoma porcino, el abordaje de estos genes no pudo realizarse mediante las

aproximaciones de análisis masivo si no de análisis estructurales y de expresión génica específicos y tradicionales.

El estudio de la secuencia del gen *LEP* en el material IBMAP ha permitido identificar 39 polimorfismos, de los cuales ocho fueron identificados por primera vez en este trabajo. De éstos, el polimorfismo intrónico *LEP*g.1387C>T mostró asociación con caracteres de crecimiento, deposición grasa y composición corporal en un análisis conjunto con el polimorfismo *LEPR*c.1987C>T, detectándose, además, efectos de ambos polimorfismos sobre la composición de ácidos grasos en grasa subcutánea, este último mediado probablemente a través de su efecto sobre la deposición grasa. Por otra parte, el análisis de la variabilidad genética de las regiones más polimórficas de ambos genes, región promotora del gen *LEPR* y la región intrónica entre los exones 2 y 3 del gen *LEP*, en diversas razas porcinas asiáticas y europeas y jabalí europeo, ha permitido identificar 16 haplotipos en el caso del gen *LEPR* y 48 en el *LEP*. Los resultados del gen *LEP* han mostrado la posible existencia de dos núcleos de domesticación del cerdo en Asia y sugiere además cierta subdivisión en los genomas de origen europeo no detectada en estudios previos, aunque este resultado requiere análisis complementarios. Además, los análisis muestran un posible evento de selección sobre el gen *LEPR* en la raza Ibérica, ya que no existe variabilidad del gen en la región analizada para ninguno de los cerdos ibéricos analizados, posiblemente asociado al especial fenotipo de esta raza en relación al apetito y la deposición grasa, ligado a su sistema tradicional de producción basado en el aprovechamiento de los recursos de la dehesa, de marcada estacionalidad. Finalmente se ha llevado a cabo un estudio de expresión de ambos genes en cinco tejidos (grasa, hígado, hipotálamo diafragma y *Longissimus dorsi*), identificándose una nueva isoforma corta del gen *LEPR* expresada principalmente en hígado, mientras que la isoforma larga muestra expresión únicamente en hipotálamo. En el caso del gen de la leptina, se ha identificado únicamente una isoforma cuya expresión más elevada se da en grasa. Además se han analizado las regiones promotoras de ambos genes, identificando 16 polimorfismos en la región promotora del gen *LEPR* y siete en la región promotora del gen *LEP*. El análisis de expresión génica condicionado por el genotipo de estos polimorfismos ha puesto de manifiesto una posible regulación específica de tejido e isoforma del gen *LEPR*. El análisis de los perfiles de expresión

de ambos genes, además de apoyar su conocida función de señalización a través de la ruta hipotalámica, parece indicar la posible implicación de ambas moléculas en funciones periféricas como el metabolismo lipídico, ya que existe expresión en tejidos distintos de los asumidos, según su conocida función.

Analizando los resultados obtenidos de forma global, las diversas aproximaciones utilizadas en la presente tesis han permitido la identificación de genes y rutas génicas que parecen desarrollar importantes funciones en el crecimiento y la deposición grasa en porcino. Un análisis de interacción *in silico* con los genes más relevantes extraídos de los distintos trabajos presentados permite visualizar relaciones e interacciones, conocidas o predichas, entre la mayoría de ellos. Este análisis pone de manifiesto una localización central y un gran número de interacciones de las proteínas codificadas por los genes *EGR1* e *IRF1* con el resto de genes analizados, conforme a la hipótesis de que se trata de importantes reguladores transcripcionales del control del crecimiento y la deposición grasa. Además de mostrar la relación de varias de las moléculas, ADIPOR2, EGR1 y NR3C1 con en el eje LEP-LEPR.

Los resultados obtenidos en la presente tesis doctoral abren nuevas e interesantes vías de estudio del crecimiento y la deposición grasa. La mayor parte de los análisis realizados han permitido detectar regiones del genoma, genes y polimorfismos interesantes, muchos de los cuales no han sido estudiados hasta la fecha en la especie porcina. Por tanto, un análisis más profundo de estos así como la validación en distintas poblaciones, permitirá identificar mutaciones causales o potentes marcadores con aplicación a la selección en poblaciones comerciales. En el caso del receptor de la leptina, la asociación del polimorfismo *LEPRc.*1987C>T con caracteres de crecimiento y deposición grasa ha sido nuevamente validada, y aunque no se tienen evidencias concluyentes de su causalidad y no se puede descartar la existencia de más de una mutación causal, los resultados apoyan el empleo de este polimorfismo como marcador útil en programas de selección.

RESUMEN REDUCIDO

El crecimiento y la deposición grasa son algunos de los caracteres económicos más relevantes en la producción porcina, además, el cerdo es utilizado como especie modelo en el estudio de la obesidad y patologías humanas con ella relacionadas, por lo que el estudio de estos caracteres suscita un gran interés. El objetivo general de esta tesis ha sido profundizar en el conocimiento de la base genética de la regulación del crecimiento y la deposición grasa en porcino, identificando genes, rutas génicas y mutaciones que afecten a estos caracteres. Este objetivo general ha sido abordado desde distintas perspectivas utilizando tanto metodologías de análisis masivos como metodologías más tradicionales.

Se han utilizado dos aproximaciones complementarias de análisis masivo del genoma sobre el material animal IBMAP para conseguir abordar el estudio de estos caracteres desde distintas perspectivas. El análisis de detección de regiones QTL explotando información de genotipado del chip de genotipado masivo de SNPs ha permitido corroborar e identificar nuevas regiones del genoma asociadas a los caracteres de interés, así como proponer potentes genes candidatos dentro de las mismas. Por otro lado, la secuenciación mediante RNA-Seq del transcriptoma hipotalámico, principal tejido regulador de la ingesta de alimentos, el balance energético y el peso corporal, ha mostrado la elevada complejidad de este tejido. Además, mediante un análisis de expresión diferencial llevado a cabo en animales divergentes para los caracteres de interés, se han podido identificar genes, tránscritos y potenciales reguladores transcripcionales involucrados en rutas relacionadas con el control del crecimiento y la deposición grasa. Así mismo, se han identificado potentes genes candidatos que se localizan dentro de las regiones QTL identificadas en el estudio anterior.

Por otra parte, se han utilizado estrategias de análisis más tradicionales para el estudio de los genes candidatos: leptina (*LEP*) y receptor de la leptina (*LEPR*). Ambos genes son especialmente relevantes en este estudio dada su importante función sobre la regulación del crecimiento y deposición grasa y los resultados previos del polimorfismo *LEPR*c.1987C>T en porcino. Además, debido a la falta de anotación del genoma porcino, el abordaje de estos genes no pudo realizarse mediante las aproximaciones de análisis masivo. El estudio del gen *LEP* en el material IBMAP ha permitido identificar un polimorfismo, *LEP*g.1387C>T, asociado

con caracteres de crecimiento, deposición grasa y composición corporal en un análisis conjunto con el polimorfismo *LEPR*c.1987C>T, detectándose, además, efectos de ambos polimorfismos sobre la composición de ácidos grasos en grasa subcutánea. Por otra parte, el análisis de la variabilidad genética de las regiones más polimórficas de ambos genes en diversas razas porcinas ha indicado, en el caso del gen *LEP*, la posible existencia de dos núcleos de domesticación del cerdo en Asia, y en el caso del *LEPR*, un posible evento de selección en la raza Ibérica posiblemente asociado al especial fenotipo de esta raza en relación al apetito y la deposición grasa. Finalmente se ha llevado a cabo un estudio de expresión de ambos genes en diversos tejidos, identificándose una nueva isoforma corta del gen *LEPR* y poniendo de manifiesto una posible regulación específica de tejido e isoforma en este gen. El análisis de sus perfiles de expresión, además de confirmar su conocida función de señalización a través de la ruta hipotalámica, parece indicar la posible implicación de ambas moléculas en funciones periféricas como el metabolismo lipídico.

En general, los resultados de la presente tesis doctoral abren nuevas vías para progresar en el conocimiento del crecimiento y la deposición grasa y permiten destacar genes, reguladores transcripcionales y mutaciones especialmente relevantes en la determinación de estos caracteres.

SUMMARY

The pork is the most worldwide consumed red meat; pig production therefore is the fastest growing livestock subsector. After China, the European Union is the largest producer, where Spain is the second producer after Germany. In this context, growth and fat deposition are some of the most important economic traits in pig production. In addition, the swine is used as a model species in the study of obesity and related diseases in human, given their greater similarity in anatomy, genetics and pathophysiology. Therefore the study of the genetic basis of growth and fatness arouses great interest from both productive and biomedical points of view

Several factors contribute to the phenotypic variation in growth and fat deposition, pointing out the complex regulation of these traits. The composition of the diet, the age and the sex have a large effect. Also, environmental factors such as production systems, the light hours and temperature influence the individual requirements affecting the fat content of the animals. Furthermore, the differences in growth and fat deposition in different pig breeds suggest the importance of genetic factors in the regulation of these processes. Previous studies has estimated the heritability of these traits in pigs at approximately 0.45 and 0.25 for fat deposition and growth traits, respectively.

The overall objective of this thesis was to deepen the knowledge of the genetic basis of the growth and fat deposition regulation in pigs, identifying genes, gene pathways and mutations affecting these traits. This general objective has been approached from different perspectives employing both high-throughput analysis and traditional methodologies, using the animal material coming from an experimental cross between Iberian x Landrace pigs, breeds showing extreme phenotypes for these traits.

The Iberian x Landrace experimental cross was developed in 1996 with the aim of studying the genetic basis of growth, fat deposition, body composition and meat quality traits in pig. First, an F2 generation was obtained and then new animal material was obtained through a F3 and two backcross generations coming from the cross of F2 males x Landrace and F1 males x Landrace pigs. The most important results obtained in this material and related to growth and fat deposition included the detection of QTLs on chromosomes SSC2, SSC4 and SSC6.

The QTL located on SSC6 has been deeper studied identifying the *LEPR* gene as biological and positional candidate to carry the mutation causal of the QTL effects. In fact, a point mutation was identified, *LEPRc*.1987C>T, strongly associated with fat deposition, body composition and growth related traits. Even more, this SNP seems to affect the level of *LEPR* gene expression in the hypothalamus, however no evidence of the causal effect of this mutation has been provided yet. The current thesis has been developed from these previous results related to growth and fat deposition.

Two complementary high-throughput analysis strategies were employed to approach the study of these traits. One analysis consisted in the detection of QTL regions by linkage analysis, exploiting the information from porcine massive genotyping SNP chip on the IBMAP F3 generation and both backcrosses. The analyses were performed using the joint information of the F3 and backcrosses and separating the two according to the male generation used to obtain the backcrosses, highlighting the differences between the expected QTL genotypes and allele random sampling of the QTLs in the F1 and F2 individuals. The results of these analyzes have allowed us to corroborate and identify new regions of the genome associated with traits of interest, even more to propose powerful candidate genes within the regions.

Specifically, we have confirmed the QTLs located in SSC4 and SSC6 for fat deposition, and their confidence intervals have been reduced from 12 and 18 cM to 7 and 2 cM, respectively, facilitating the identification of candidate genes. Additionally, nine new QTLs were identified on chromosomes SSC1, SSC2, SSC4, SSC5, SSC9, SSC11, SSC13, SSC14 and SSC17, associated with different growth and fat deposition traits, identifying for most of them powerful functional and positional candidate genes as *PTPRD* genes, *VCAM1*, *AGL*, *ADIPOR2*, *VDR*, *MYOG*, *FASLG*, *NCK1*, *CYP2E1* and *ID1*.

Furthermore, the sequencing of the porcine hypothalamic transciptome by RNA-Seq technology was carried out in animals divergent for growth and fat deposition coming from one of the IBMAP backcrosses. The hypothalamus is the main regulatory centre of food intake, energy balance and body weight. In this paper, the high complexity of the hypothalamic transcriptome in comparison with

others such as hepatic and gonads one, and its high implication in the regulation of the traits of interest, have been reported.

In addition, a differential expression analysis allowed us to detect more than 200 known transcripts and 50 novel isoforms differentially expressed between the two groups of divergent growth and fat deposition pigs. The biological and functional annotation has allowed us to identify genes, transcripts and potential transcriptional regulators involved in pathways related to growth control and fat deposition, several of them validated by quantitative PCR.

Likewise, powerful candidate genes have been identified within QTL regions identified in the previous study, especially for chromosomes 2, 4, 5, 6, 14 and 17. In summary, a set of 21 relevant differentially expressed genes, functionally related to the traits of interest could be highlighted. A search for the potential regulators of this set of genes revealed six transcription factors located within the QTL regions previously detected, potentially regulating some of these genes: *IRF1, EGR1, PBX1a, POU2F1, NR3C1* and *NF-Yb.* A functional network was constructed with the relevant genes and transcription factors to analyzed potential interactions and links between them. An overview of the different approaches used, highlight the *EGR1* and *IRF1* genes, because they are both differentially expressed, positional candidates, located in the QTL SSC2, as well as functional candidates. The *IRF1* gene encodes a transcription factor involved in the regulation of growth hormone. Likewise the *EGR1* gene encodes a transcription factor that plays an important role in the regulation of food intake and in the duration and amount of intake. In addition, even it does not appear differentially expressed in this study, *NR3C1* gene is also considered relevant because it is a transcription factor that regulates expression of five genes identified as differentially expressed in the analysis and has a central location in the network built.

In addition, more traditional analysis strategies have been applied for the study of the candidate genes: *leptin* (*LEP*) and *leptin receptor* (*LEPR*). Both genes are particularly relevant in this study due to its important role on the regulation of appetite, growth and fat deposition, and the previous results obtained for the *LEPRc*.1987C>T polymorphism in pigs. The leptin is a peripheral hormone regulating satiety and produced primarily by white adipocytes conditional on the

amount of fat. It is released into the bloodstream through which reaches its specific receptor (LEPR) located mainly in the hypothalamus and producing its activation and triggering a signaling cascade of transcription factors that induces satiety and energy expenditure. Due to the lack of porcine genome annotation, the approach to study these genes could not be performed by massive analysis approximations; therefore they have been analyzed using specific and more traditional structural and gene expression analyses.

The study of the *LEP* gene sequence in the material IBMAP allowed us to identify 39 polymorphisms, eight identified in this study for the first time. Among them, the intronic polymorphism *LEPg*.1387C>T showed association with growth, fat deposition and body composition traits in a joined analysis with the *LEPRc*.1987C>T polymorphism. Also effects of both polymorphisms on fatty acid composition in subcutaneous fat was detected, the latter probably mediated through its effect on fat deposition.

Moreover, the analysis of the genetic variability of the most polymorphic regions of both genes, promoter region of the *LEPR* gene and intron region between exons 2 and 3 of the *LEP* gene, in several pig breeds from Asian and Europe and European wild boars, allowed the detection a total of 16 haplotypes for the *LEPR* gene and 48 for the *LEP*. The results of the LEP gene shown the possible existence of two centers of pig domestication in Asia and further suggests a certain subdivision in the genomes of European origin undetected in previous studies, although latter result requires additional analysis. In addition, the analyses revealed a potential selection event on the *LEPR* gene in the Iberian breed, since there is not gene variation into the target region for either of the analyzed Iberian pigs, possibly associated with the particular phenotype of this breed in relation to appetite and fat deposition, linked to their traditional production system based on the use of seasonal pasture resources.

Finally, a study of gene expression patterns was conducted for both *LEP* and *LEPR* genes in five tissues (hypothalamus, backfat, liver, diaphragm and *Longissimus dorsi*). A new *LEPR* short isoform (*LEPRa*) expressed primarily in the liver was identified, while the long isoform (*LEPRb*) displayed expression almost exclusively in hypothalamus. Instead only one *LEP* isoform was identified, showing

its highest expression in fat tissue, as expected. The promoter regions of both genes have also been analyzed, identifying 16 polymorphisms in the promoter region of the *LEPR* gene and seven in the *LEP* gene promoter region. The analysis of gene expression conditioned on the genotype of these polymorphisms revealed a potential tissue and isoform specific regulation of the *LEPR* gene. The analysis of the expression profiles of both genes, besides supporting their known role in signaling function through the hypothalamic pathway, supports the potential involvement of both molecules in peripheral functions such as lipid metabolism, since there are high expression levels in tissues other than those assumed due to their known function.

Analyzing the whole obtained results, the various approaches used in this thesis have allowed us to identify genes and gene pathways that seem to develop important roles in growth and fat deposition in pigs. An *in silico* interaction analysis with the core genes extracted from the presented studies allowed us to visualize the relationships and interactions, known or predicted, among most of them. This analysis reveals a central location and a great number of interactions among the proteins encoded by the *EGR1* and *IRF1* genes and the other included genes, according to the hypothesis that both factors are important transcriptional regulators of growth control and deposition fat. Besides it shows the relation between several molecules such as ADIPOR2, EGR1 and NR3C1 with the LEP- LEPR axis.

The results obtained in this thesis open up exciting new avenues for studying the genetic regulation of growth and fat deposition. Most of the analyses have identified interesting regions of the genome, genes and polymorphisms, many of which have not been studied to date in pig. Therefore, further analysis of these and validation in different populations, will probably allow us to identify causal mutations or powerful genetic markers with application to the selection in commercial populations.

BRIEF SUMMARY

Growth and fatness are some of the most important economic traits in swine production. Moreover, the pig has been used as a model species for the study of obesity and related diseases in human, therefore the study of these traits is of great interest. The overall objective of this thesis was to deepen the understanding of the genetic basis of the regulation of growth and fatness in pigs, identifying genes, gene pathways and mutations affecting these traits. This general objective has been approached from different perspectives using high-throughput analyses and traditional methodologies.

Two complementary high-throughput approaches have been used on IBMAP experimental cross animal material to study growth and fatness traits from different perspectives. A QTL detection analysis was performed exploiting the genotyping information from the high-density porcine SNP chip to corroborate and identify genomic regions associated with the traits of interest and identify potent candidate genes underlying these QTL regions. Furthermore, the analysis carried out by RNA–seq to sequence the hypothalamic transcriptome, main tissue regulating food intake, energy balance and body weight, showed the high complexity of this tissue. In addition, a differential expression analysis carried out in animals divergent for the traits of interest, has allowed us to identify genes, transcripts and potential transcriptional regulators of pathways involved in growth and fatness control. Likewise, powerful candidate genes located within the QTL regions identified in the previous study have been proposed.

Traditional analytical strategies have also been used for the study of the candidate genes leptin (LEP) and leptin receptor (LEPR). Both genes are particularly relevant in this study because of their important role on the regulation of growth and fatness and previous results of the porcine LEPRc.1987C>T polymorphism. Moreover, due to the lack of porcine genome annotation, the high-throughput approach could not be used to study these genes. The LEP gene study conducted in the IBMAP animal material has allowed us to identify a polymorphism, LEPg.1387C>T, associated with growth, fat deposition and body composition traits, in a joint analysis with the polymorphism LEPRc.1987C>T. Additionally, effects of both polymorphisms on the fatty acid composition of

subcutaneous fat were also detected. Moreover, the analysis of the genetic variability of the most polymorphic regions of both genes in different pig breeds

has allowed us to detect in the case of the LEP gene, the possible existence of two centers of pig domestication in Asia, in agreement with some previous results, and in the case of LEPR, a possible selection event in the Iberian breed probably associated with the especial phenotype of this breed for appetite and fat deposition. Finally, an expression study has been carried out for both genes in various porcine tissues, identifying a new LEPR short isoform and revealing a potential isoform and tissue specific regulation of this gene. The analysis LEP and LEPR expression profiles, confirmed its known role in the hypothalamic signaling pathway, and suggested the possible implication of both molecules in peripheral functions such as lipid metabolism.

In general, the results of this thesis opens new tracks to progress in the study of growth and fatness and highlights genes, transcriptional regulators and mutations particularly relevant in the determination of these traits.

INTRODUCCIÓN

Sus scrofa

El cerdo (*Sus scrofa*) es un mamífero perteneciente a la familia *Suidae* y al género *Sus*. Evidencias arqueológicas y genéticas sugieren que fue originado a partir de jabalíes por múltiples procesos independientes de domesticación que tuvieron lugar principalmente en Asia y la antigua Mesopotamia (el creciente fértil) y posteriormente en Europa y otras regiones, y datan alrededor de 700.000 años la divergencia entre los ancestros de los cerdos originarios asiáticos y europeos (Larson et al. 2005, Groenen et al. 2012). Alrededor de 9.000 años después de su domesticación, el encuentro entre cerdos de razas asiáticas y europeas se produjo, según estudios previos, hace alrededor de 300 años cuando los cerdos asiáticos se importaron a Europa (White 2011). Estos milenios de divergencia genética y cultural, junto con el hecho de estar sometidos a diferente presión evolutiva hicieron que ambas poblaciones desarrollaran fenotipos marcadamente diferentes. El primer cambio relevante en la producción de cerdos se produjo en Inglaterra durante los siglos XVII y XVIII en respuesta a la creciente demanda de productos cárnicos (White 2011, Porter 1993). Fue en este momento cuando las razas asiáticas llegaron a Inglaterra, jugando un papel fundamental en la transformación de las razas inglesas. El proceso culminó a finales del siglo XVIII con la generación de cruces que combinaban el gran tamaño y crecimiento tardío de los cerdos del norte de Europa con la mayor precocidad inicial (crecimiento temprano) y mayor contenido graso de las razas asiáticas (White 2011). A partir de entonces, las diversas razas porcinas fueron seleccionadas de acuerdo con criterios específicos de diferentes grupos de criadores, originando un gran número de razas porcinas especializadas, con fenotipos muy diferentes. Procesos similares realizados en otros países dieron lugar a un gran número de razas diferentes. En la actualidad, pese a la notable disminución experimentada en las últimas décadas, 350 razas de cerdo están recogidas en la página web del Instituto Internacional de Investigación Ganadera (http://agtr.ilri.cgiar.org/index.php?option=com_content&view=article&id=240&Itemid=298). La mayoría de estas razas modernas son razas locales, mientras que solo unas pocas razas intensamente seleccionadas se utilizan internacionalmente en la industria porcina (Ollivier et al. 2009).

Importancia de la especie porcina

La industria porcina

De acuerdo con la FAO, la carne roja de mayor consumo mundial es la carne de cerdo, cuya demanda en las últimas décadas ha experimentado un fuerte incremento debido al consumo de productos derivados del cerdo en países en desarrollo. Junto con el de las aves de corral, el porcino es el subsector pecuario de mayor crecimiento y se prevé que alcanzará los mil millones de animales antes de 2015. La producción porcina está distribuida por todo el mundo, con exclusión de algunas regiones que mantienen reservas culturales o religiosas en relación con el consumo de carne de cerdo.

De acuerdo a los datos del Ministerio Español de Agricultura Alimentación y Medio Ambiente, China es el mayor productor de carne de cerdo a nivel mundial, generando alrededor de un 50% de la producción porcina total (Figura 1), seguida de la Unión Europea con un 21% de la producción total.

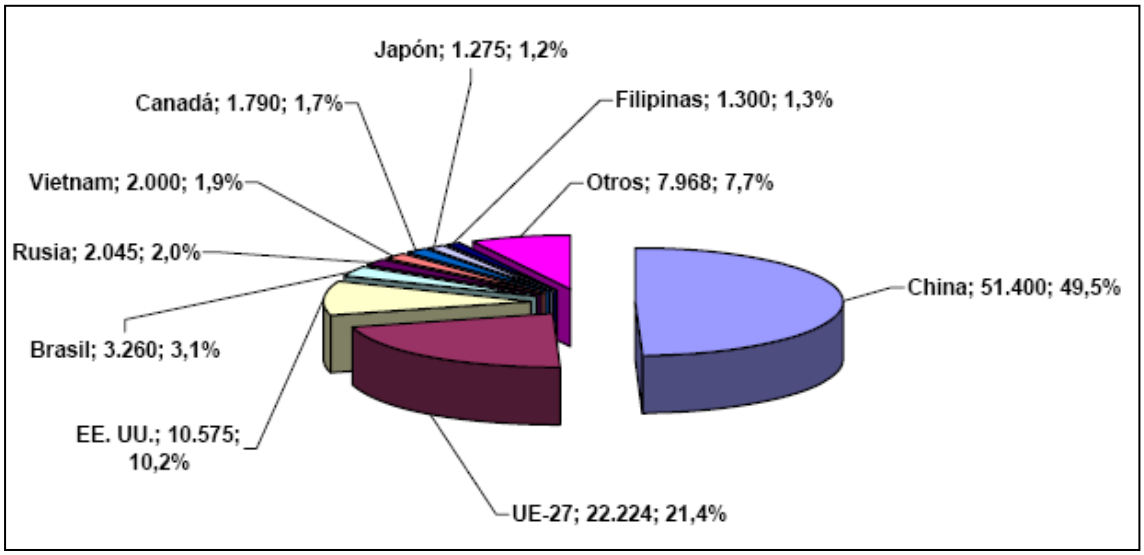


Figura 1. Principales países productores de carne de cerdo en el año 2012 (miles de toneladas)(MAGRAMA).

Dentro de la Unión Europea, España es el segundo mayor productor de carne de cerdo, después de Alemania. La producción porcina española supone aproximadamente un 16% de la producción total de la Unión Europea (Figura 2).

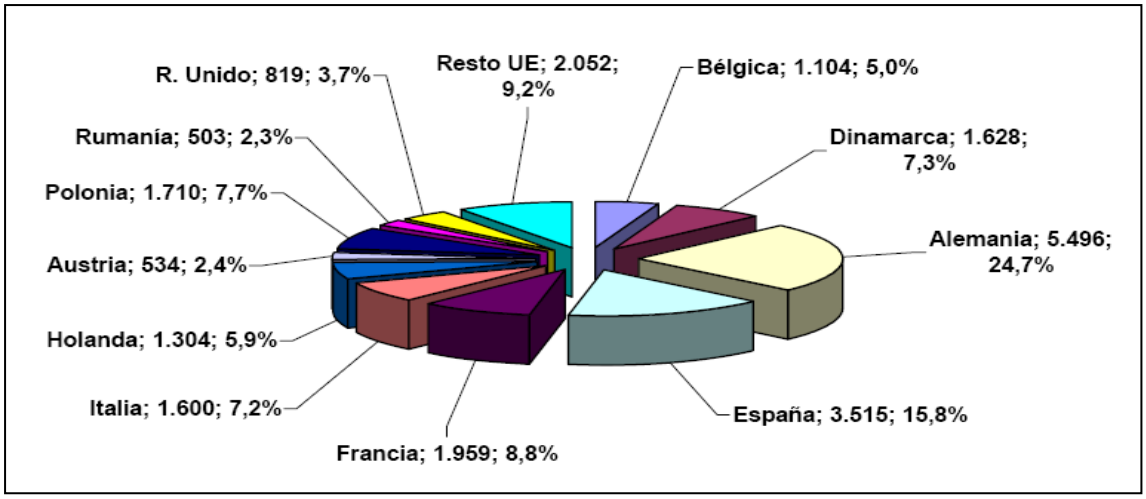


Figura 2. Producción de carne de cerdo en la unión europea durante el año 2012 (miles de toneladas)(MAGRAMA).

Cabe mencionar además el gran incremento que han experimentado las exportaciones españolas de productos porcinos en los últimos años, ya sea hacia otros países de la Unión Europea o hacía el resto de países. En los últimos diez años, se han triplicado las exportaciones españolas de productos porcinos. En el pasado año 2012 alrededor de 1.400.000 tonelas de carne de cerdo fueron exportadas a otros países (Figura 3).

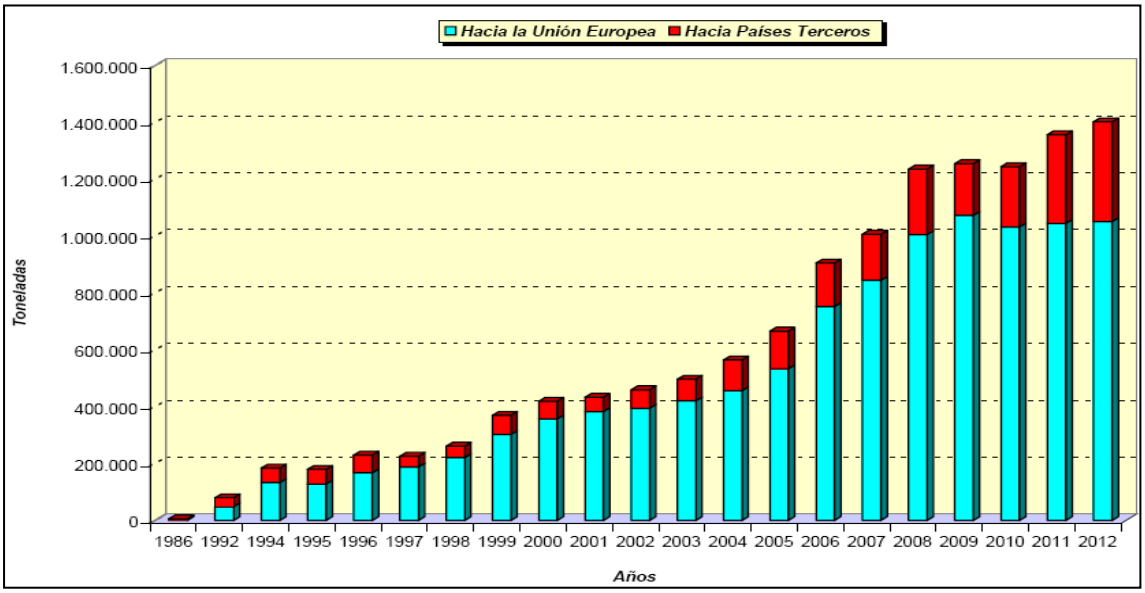


Figura 3. Evolución de las exportaciones españolas en el sector de la carne porcina (toneladas)(MAGRAMA).

Todo ello pone de manifiesto la gran importancia del sector porcino en España y el interés del estudio de los diferentes aspectos (genéticos, nutricionales, etc.) influyentes en los caracteres productivos en esta especie.

El cerdo como especie modelo

El cerdo viene siendo utilizado como especie modelo desde hace algunas décadas, proporcionando un mayor conocimiento de la progresión de enfermedades y nuevos tratamientos potenciales. A pesar de que los organismos modelo clásicos, rata y ratón, han proporcionado una gran cantidad de información sobre la biología básica de genes y proteínas, la utilidad de estos modelos es limitada por su incapacidad de representar suficientemente bien las enfermedades humanas de forma fidedigna. Estos animales modelo no reflejan adecuadamente las patologías de muchas enfermedades humanas. Sin embargo, el cerdo ha sido una especie muy utilizada como modelo para la salud humana y las enfermedades, dada su mayor semejanza en anatomía, genética y fisiopatología. Además, comparado con otras especies modelo de mayor tamaño, presenta ventajas como una temprana madurez sexual (a los 6-8 meses), una capacidad reproductiva durante todo el año, un corto periodo de gestación (de aproximadamente cuatro meses) y partos múltiples (Walters et al. 2012).

La publicación del genoma porcino ha sido un hito clave que ha permitido analizar los mecanismos moleculares de diversas enfermedades humanas, mejorando la identificación de genes relevantes para estas enfermedades. El genoma del cerdo es aproximadamente un 7% más pequeño que el humano, mientras que los de ratón y perro son un 14% más pequeños. Además de la similitud en su tamaño, ambos genomas presentan una extensa homología. A nivel de nucleótido, la especie porcina es tres veces más similar a la humana que la especie murina (Cabot et al. 2001); asimismo, existen grandes bloques sinténicos entre ambas especies, lo que hace que las estrategias de clonaje posicional puedan aplicarse directamente en humano y que un gran número de las interacciones reguladoras estén conservadas. Debido a todo ello, la genómica comparativa permite identificar e investigar genes, elementos reguladores y polimorfismos, conservados entre ambas especies, potencialmente asociados a enfermedades (Walters et al. 2012).

El crecimiento y la deposición grasa en porcino

El crecimiento y la deposición grasa son algunos de los caracteres económicos más importantes en relación a la producción porcina, por lo que generan gran interés y han sido ampliamente estudiados. Diversos factores contribuyen a la variación fenotípica en estos caracteres. La composición de la dieta, la edad y el sexo tienen un gran efecto sobre la deposición grasa y la tasa de crecimiento. Así mismo, factores medioambientales como los sistemas de producción, las horas de luz y la temperatura influencian los requerimientos individuales afectando al contenido graso de los animales. Por otro lado, las diferencias existentes en el crecimiento y la deposición grasa de las distintas razas porcinas indican la importancia de los factores genéticos en la regulación de estos procesos (Ai et al. 2011). La heredabilidad de estos caracteres en porcino se ha estimado en aproximadamente 0.45 para la deposición grasa y 0.25 para caracteres de crecimiento (Hetzer and Harvey 1967, Siers and Homson 1972).

Control del apetito

A pesar del gran número de regiones del cerebro involucradas en la regulación de la ingesta de alimentos, el hipotálamo se considera el principal centro regulador del apetito, la ingesta alimentaria y del peso corporal (Hillebrand et al. 2002). El hipotálamo es crucial en la detección del nivel de nutrientes y en la modulación de la ingesta de alimentos y del gasto energético en respuesta a señales periféricas provenientes de otros tejidos (Simpson et al. 2009). El hipotálamo es una región nuclear, es decir, está formado por una serie de núcleos interconectados entre sí: el núcleo arcuato (ARC), el paraventricular (PVN), el área hipotalámica lateral (LHA), el núcleo ventromedial (VMH) y el núcleo dorsomedial (DMH). El ARC es el núcleo principal, está estratégicamente localizado para integrar las señales provenientes de la periferia encargadas de regular el apetito y posee dos tipos de neuronas encargadas de controlar el balance energético. El primer grupo de neuronas son las que sintetizan neuropéptidos orexigénicos como el neuropéptido Y (NPY) y la proteína Agouti (AgRP). El segundo grupo está formado por neuronas que producen neuropéptidos anorexigénicos o supresores del apetito como la pro-opiomelanocortina (POMC) o el transcrito relacionado con la cocaína-amfetamina (CART). Desde el núcleo ARC las señales se proyectan hasta los núcleos

secundarios PVN, LHA, VMH y DMH que se organizan formando una compleja red de señales neuronales que responden a la señal de saciedad y cuya comunicación resulta esencial para la regulación de la homeostasis energética a largo plazo.

Existen numerosas moléculas producidas fuera del sistema nervioso central que intervienen en la regulación del apetito y que actúan sobre las células del sistema nervioso, sobre neuronas cerebrales directamente o a través del nervio vago, que constituye un importante enlace entre el tracto gastrointestinal y el cerebro. Las adipoquinas son secretadas por el tejido adiposo y las más importantes son la leptina, la adiponectina y la resistina. La leptina es una hormona periférica reguladora de la saciedad producida principalmente por los adipocitos en función de la cantidad de materia grasa. Es liberada al torrente sanguíneo a través del cual llega a su receptor específico (LEPR) localizado principalmente en el hipotálamo produciendo su activación y desencadenando una cascada de señalización que produce sensación de saciedad. La adiponectina es otra hormona que aumenta la sensibilidad a la insulina participando en el metabolismo de la glucosa y los ácidos grasos. La insulina es otra molécula señalizadora de la adiposidad que es secretada por las células pancreáticas en función del nivel de glucosa y que es el principal activador del almacenamiento energético en el tejido adiposo (Keiffer and Habener, 2000).

Por otro lado, el tracto gastrointestinal también libera una serie de hormonas peptídicas en función del nivel de nutrientes, que regulan el hambre y la sensación de saciedad. Entre ellas cabe destacar la colecistokinina (CCK), el péptido similar al glucagón (GLP-1), el polipéptido pancreático (PP) y el péptido YY (PYY), que se liberan a la circulación tras la comida en proporción a las calorías consumidas y actuan a través del nervio vago (Turton et al. 1996) y la oxintomodulina, todas ellas inhibidoras del apetito. Por otro lado la grelina, producida por el estómago, es la encargada de generar la sensación de hambre anterior a una comida teniendo un potente efecto estimulador sobre la ingesta de alimento (Shrestha et al. 2004).

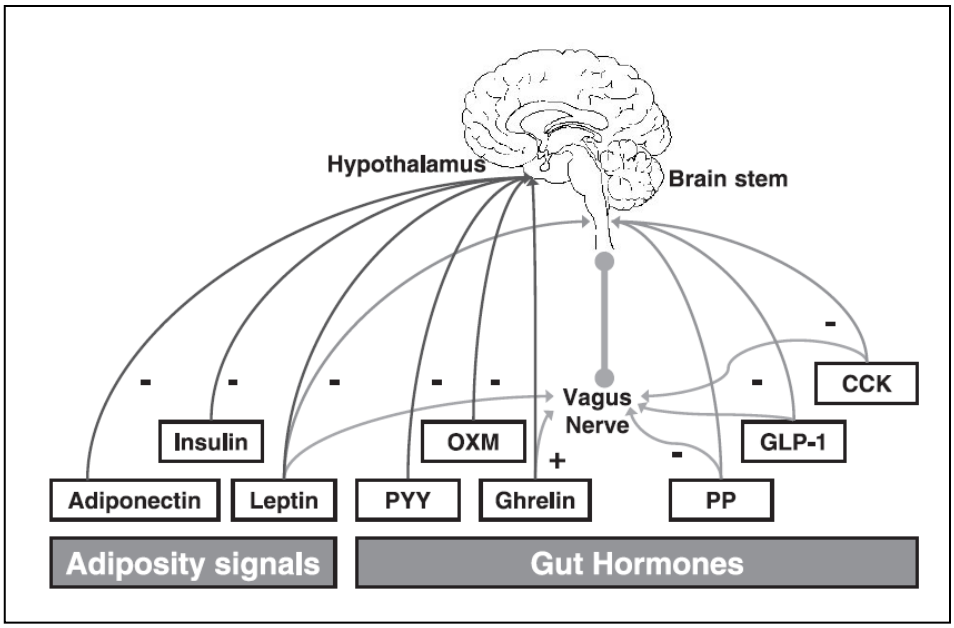


Figura 4. Esquema de la regulación del apetito y la homeostasis energética por las señales periféricas del tejido adiposo y del tracto gastrointestinal que regulan los circuitos centrales del hipotálamo y el cerebro (Stanley et al. 2005).

Metodologías de análisis molecular aplicadas a la mejora genética animal

Una gran variedad de herramientas moleculares y métodos estadísticos han sido empleados en el área de la mejora genética animal con el fin de conocer la base genética de la regulación de caracteres complejos como son el crecimiento, la deposición grasa, la prolificidad, etc.

Durante los años 90 empezaron a realizarse numerosos estudios centrados en las variaciones en la secuencia del ADN, especialmente con marcadores de tipo microsatélite, con el fin de construir mapas genéticos y detectar regiones del genoma o quantitative trait loci (QTL) asociados a determinados caracteres (Ollivier et al. 2009). El primer estudio de este tipo en animales domésticos fue llevado a cabo por Andersson et al. (1994) en un cruce de jabalí x Large White centrado en la detección de QTLs asociados a caracteres de crecimiento y deposición grasa. Posteriormente, se han desarrollado un gran número de estudios para confirmar o detectar nuevos QTLs asociados a los caracteres de interés, recogidos en la base de datos pública de QTLs de cerdo (PigQTLdb) (Hu et al. 2007, Hu et al. 2013). En la actualidad (3 de septiembre de 2013) hay un total de 8.935

QTLs en esta base de datos, asociados a un total de 644 caracteres diferentes descritos en 371 publicaciones (Tabla 1).

Tabla 1. Número de QTLs identificados hasta el momento (3 de septiembre de 2013) en el genoma porcino por clases de caracteres según PigQTLdb.

Clase de caracteres	Número de QTLs identificados
Calidad de la carne y la canal	5.755
Salud	883
Reproducción	803
Rasgos exteriores	786
Producción	708
TOTAL	**8.935**

El diseño experimental más utilizado en este tipo de estudios para la detección de QTLs son las poblaciones F2 generadas a partir del cruce de líneas divergentes para los caracteres a analizar, en las que se suponen alelos alternativos para los QTLs (Choi et al. 2010). Mediante esta estrategia se identificó un considerable número de QTLs asociados a caracteres de interés. Una limitación del uso de marcadores de tipo microsatélite, ampliamente utilizados en estos estudios, es la excesiva distancia entre los distintos marcadores, que hace que los QTLs identificados tengan unos intervalos de confianza de decenas de cM, en los que es difícil identificar de forma exitosa genes candidato y las potenciales mutaciones causales de los efectos detectados.

En el año 2003 se estableció el consorcio para la secuenciación del genoma porcino (Swine Genome Sequencing Consortium, SGSC) con el objetivo de llevar a cabo la secuenciación y caracterización del genoma porcino, compuesto por 18 pares de autosomas y dos cromosomas sexuales (cromosomas X e Y) y con un tamaño estimado de alrededor de 2.7 Gb (Walters et al. 2012). En primer lugar se llevó a cabo la secuenciación de una hembra de la raza Duroc usando cromosomas artificiales de bacterias (BACs) obteniendo una cobertura de 4x (Archibald et al. 2010). Desde entonces, se han puesto a disposición de la comunidad investigadora diversas versiones del genoma porcino que se han revisado con el objetivo de

mejorarlas al máximo hasta la actual versión Sscrofa 10.2 (Groenen et al. 2012). La disponibilidad del genoma porcino ha proporcionado cientos de miles de nuevos marcadores genéticos permitiendo un gran avance en la detección de regiones genómicas asociadas a los caracteres de interés.

Las limitaciones que tiene el uso de marcadores microsatélites ha sido superada con el uso de marcadores de tipo SNP y la aparición de los chips de genotipado masivo. A pesar de que los SNPs son marcadores bialélicos y por tanto menos informativos que los microsatélites, su abundancia y automatización en el genotipado han convertido a estos marcadores en los más populares. Hoy en día se han desarrollado chips de genotipado masivo de SNPs para la mayoría de especies ganaderas, incluido el porcino, que permiten el genotipado de miles de SNPs distribuidos uniformemente a lo largo de todo el genoma (Fan et al. 2010). El uso de un panel de marcadores tan denso permite no solamente refinar las posiciones de los QTLs previamente identificados, sino además identificar nuevos QTLs en regiones no bien cubiertas en estudios previos y obtener intervalos de confianza más reducidos facilitando la búsqueda de genes y mutaciones causales. En concreto, el chip comercial disponible para la especie porcina PorcineSNP60, fue desarrollado por Ramos et al. (2009) usando las cuatro razas porcinas más comúnmente utilizadas en producción (Duroc, Pietrain, Landrace y Large White) y el jabalí, lo que hace que sea una herramienta muy útil para la mayoría de estudios genéticos desarrollados en porcino. Este chip contiene un total de 64.232 SNPs, de los cuales 58.821 están mapeados en la versión Sscrofa 10.2 del genoma porcino (Ramos et al. 2009).

La selección genómica es una de las aplicaciones que más repercusión ha tenido como resultado directo del uso de los chips de genotipado masivo; se trata de una forma avanzada de selección asistida por marcadores que utiliza la información de un gran número de marcadores distribuidos por todo el genoma y aprovecha el desequilibrio de ligamiento de las poblaciones ganaderas (Fan et al. 2010). Meuwissen et al. (2001) definieron la selección genómica como una estrategia de predicción de valores mejorantes de los animales utilizando información genética proveniente de un denso panel de SNPs que cubren gran parte del genoma. Así, los datos de genotipos y fenotipos de una población de referencia se utilizan para

estimar los efectos de cada marcador y generar una ecuación capaz de predecir el valor mejorante de cada animal. Posteriormente esta ecuación predictiva puede ser aplicada a otro grupo de animales que poseen información genotípica pero no fenotípica para estimar los valores mejorantes y seleccionar los mejores animales para ser cruzados (Figura 5).

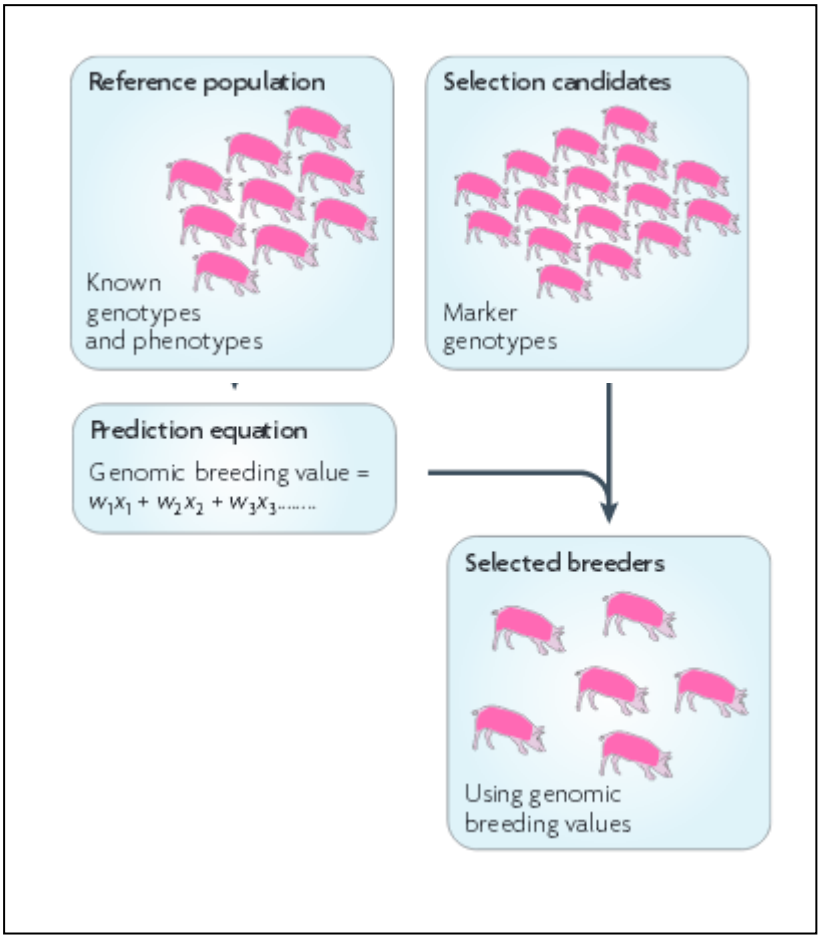


Figura 5. Esquema de la estrategia general de selección genómica (Goddard and Hayes 2009).

Otra de las principales aplicaciones de los chips de genotipado masivo de SNPs son los estudios de asociación genómica, también conocidos como GWAS por sus siglas en inglés, Genome-Wide Associations Studies. En un estudio de asociación genómica se relacionan datos de registros fenotípicos de caracteres de interés, con los datos de genotipado de un denso panel de marcadores representativo de todo el genoma, para detectar asociaciones entre los caracteres de interés y los marcadores analizados (Goddard and Hayes 2009). Los principales éxitos de esta

estrategia se han obtenido para caracteres monogénicos y oligogénicos. Para caracteres complejos los resultados de GWAS suelen indicar un gran número de polimorfismos asociados, sugiriendo que la mayoría de las mutaciones individuales tienen un efecto pequeño. En cualquier caso, la confirmación de los marcadores con efectos significativos en estudios independientes suele ser escasa. Esto se debe principalmente, a que la magnitud del efecto suele ser pequeña (lo que requeriría un experimento con un gran número de individuos que permitiera alcanzar la potencia necesaria para confirmar tales efectos), a la existencia de desequilibrio de ligamiento entre el SNP y el QTL en la población analizada pero no en todas las poblaciones y al alto número de falsos positivos que suelen obtenerse en este tipo de estudios a pesar de las correcciones aplicadas (Goddard and Hayes 2009). El método ha sido aplicado a la investigación de la base genética de caracteres de interés deficientemente conocidos, como los niveles de androsterona y estradiol responsables del olor sexual en machos enteros (Duijvesteijn et al. 2010) o parámetros de respuesta inmune en cerdos (Lu et al. 2013).

La secuenciación del genoma porcino ha permitido también el desarrollo y aplicación de las tecnologías de análisis de expresión génica globales en esta especie. En primer lugar se desarrollaron los microarrays, que permitían analizar la expresión de miles de genes conocidos. En concreto para porcino se han desarrollado dos chips comerciales, el chip de la plataforma *Affymetrix Porcine GeneChip™,* que contiene 23.937 conjuntos de sondas correspondientes a 23.256 tránscritos de 20.201 genes conocidos y el *Agilent Porcine Gene Expression Microarray* que contiene 43.803 sondas. Sin embargo, en los últimos años la aparición de las tecnologías de secuenciación masiva han revolucionado las técnicas de análisis global de la expresión génica. La secuenciación masiva del transcriptoma realizada mediante la técnica conocida como RNA-seq permite ilustrar la gran complejidad del transcriptoma generando una visión global sin precedentes permitiendo un análisis mucho más exhaustivo (Chen et al. 2011). Esta estrategia presenta varias ventajas frente al uso de microarrays. En primer lugar, tiene mayor sensibilidad y rango dinámico y menor variación técnica y ruido, además requiere menor cantidad de ARN de partida (Oshlack et al. 2010, Chen et al. 2011). En segundo lugar, mediante RNA-seq es posible capturar casi todos los tránscritos expresados, mientras que los análisis basados en microarrays

dependen de información *a priori* y no son capaces de detectar nuevos genes o tránscritos.

El procedimiento general en este tipo de análisis consiste en la fragmentación del ARN y secuenciación de fragmentos cortos mediante alguna de las tecnologías disponibles en el mercado como la de Illumina, SOLiD o Roche generando secuencias de entre 35 y 500pb (Martin and Wang 2011). Posteriormente estas lecturas o secuencias cortas son mapeadas frente al genoma de referencia, en el caso de estar disponible, o alineadas *de novo*. En un tercer paso, las lecturas son ensambladas en fragmentos más largos dentro de los genes o tránscritos. El análisis del nivel de expresión de cada tránscrito puede realizarse debido a que el número de lecturas obtenidas es proporcional al nivel de expresión. Así, una vez normalizados los datos es posible obtener los niveles de expresión de cada tránscrito y realizar análisis de expresión diferencial (Oshlack et al. 2010).

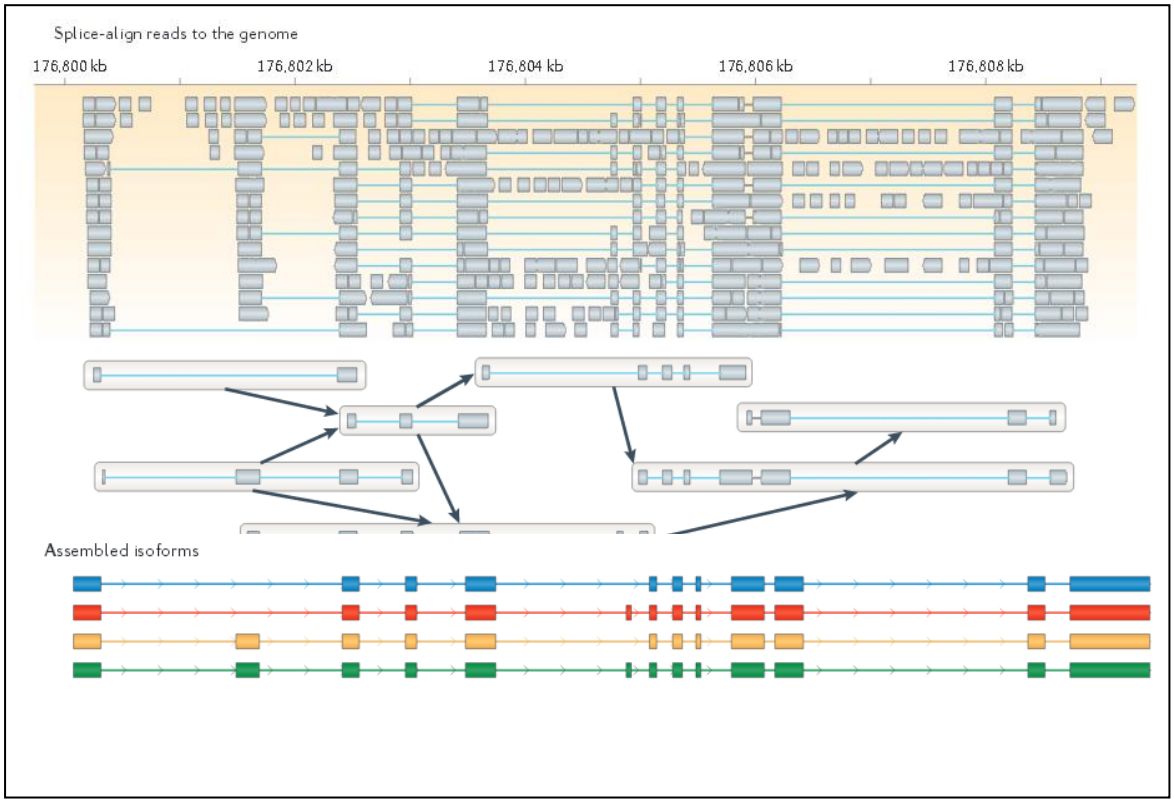


Figura 6: Representación del proceso general de ensamblado del transcriptoma basado en un genoma de referencia (Martin and Wang 2011).

El objetivo de un estudio de expresión diferencial consiste en identificar los genes cuya expresión ha cambiado significativamente entre dos condiciones diferentes. Es posible identificar no solo genes diferencialmente expresados sino también isoformas diferencialmente expresadas, diferente uso de promotores y diferentes sitios de inicio de la transcripción (Trapnell et al. 2012). Por otro lado, el análisis de RNA-seq permite un gran número de aplicaciones además de analizar la cantidad de expresión y la detección de genes diferencialmente expresados entre condiciones, como la posibilidad de inferir splicing alternativo (Sultan et al. 2008, Gan et al. 2010), detectar fusión génica (Maher et al. 2009, Pflueger et al. 2011), identificar RNAs largos no codificantes (Guttman et al. 2010) e identificar polimorfismos en regiones expresadas (Chepelev et al. 2009). Las aplicaciones de la tecnología de RNA-seq son muy diversas, como también lo son en la actualidad los métodos computacionales de análisis de este tipo de datos (Tabla 2).

Tabla 2. Algunos de los software existentes para un análisis de datos centrado en el estudio de expresión diferencial (Oshlack et al. 2010)

Analysis step	Method	Implementation
Mapping	General aligner	GMAP/GSNAP
		BFAST
		BOWTIE
		CloudBurst
		GNUmap
		MAQ/BWA
		PerM
		RazerS
		Mrfast/mrsfast
		SOAP/SOAP2
		SHRiMP
	De novo annotator	QPALMA/GenomeMapper/PALMapper
		SpliceMap
		SOAPals
		G-Mo.R-Se
		TopHat
		SplitSeek
	De novo transcript assembler	Oases
		MIRA
Summarization	Isoform-based	Cufflinks
		ALEXA-seq
	Gene-based	Count exons only
		Exon junction libraries
Normalization	Library size	
	RPKM	ERANGE
	TMM	edgeR
	Upper quartile	Myrna
Differential expression	Poisson GLM	DEGseq
		Myrna
	Negative binomial	edgeR
		DESeq
		baySeq
Systems biology	Gene Ontology analysis	GOseq

A pesar de las grandes ventajas que presenta esta tecnología, es necesario tener en cuenta que también presenta algunas limitaciones, desde las tecnologías de secuenciación propiamente hasta el análisis bioinformático de los datos. La mayoría de los sesgos en la construcción de las librerías y en la secuenciación quedan solventados con el uso de lecturas pareadas. Sin embargo, es necesario seguir implementando métodos de análisis y almacenamiento de la enorme cantidad de información que se obtiene con un experimento de RNA-seq (Chen et al. 2011).

Antecedentes del presente estudio

El primer paso para la identificación de genes responsables del crecimiento y la deposición grasa fue la detección de regiones del genoma o quantitative trait loci (QTLs) asociados a estos caracteres. El primer estudio de este tipo, mencionado anteriormente, fue llevado a cabo por Andersson et al. (1994) en un cruce de jabalí x Large White en el que se detectaron QTLs asociados a crecimiento y deposición grasa en los cromosomas 4 y 13. Posteriormente, diversos estudios se han desarrollado para confirmar o detectar nuevos QTLs asociados a estos caracteres detectándose QTLs en la mayoría de los cromosomas porcinos (Marklund et al. 1999, Walling et al. 2000, Bidanel et al. 2001, Quintanilla et al. 2002, Liu et al. 2007, Choi et al. 2010, Ai et al. 2011). En la actualidad se estudian más de 70 caracteres diferentes relacionados con el crecimiento y la deposición grasa. En total se han descrito 1.461 QTLs asociados a la deposición grasa y 598 QTLS asociados al crecimiento según la base de datos de QTLs de porcino (PigQTLdb; http://www.animalgenome.org/cgi-bin/QTLdb/SS/summary?summ=type&qtl=8,935&pub=371&trait=644), que se distribuyen en todos los cromosomas.

Además se han llevado a cabo diversos estudios con el objetivo de estudiar genes candidatos posicionales para estos caracteres como el *IGF2* (Van Laere et al. 2003), *PPARD* (Wang et al. 2003), *HMGA1* (Kim et al. 2006), *MC4R* (Fan et al. 2009) y *FTO* (Fontanesi et al. 2009, Fan et al. 2009). Sin embargo, aún queda un largo camino para descifrar la mayoría de las variantes genéticas responsables de las variaciones en el crecimiento y la deposición grasa (Ai et al. 2011).

El cruce experimental Ibérico x Landrace (IBMAP) se desarrolló en 1996, mediante la colaboración de grupos de investigadores del INIA, IRTA y UAB, con el objetivo general de estudiar la base genética de caracteres relacionados con el crecimiento, la composición corporal, la deposición grasa y la calidad de la carne. La población fue generada a partir del cruce de tres machos Ibéricos de la línea Guadyerbas (Dehesón del Encinar, Toledo) con 30 hembras Landrace (Novagenética, Solsona, Lleida). El cerdo Ibérico presenta un fenotipo especial en relación con el crecimiento y la deposición grasa. Esta raza se caracteriza por tener una composición corporal extremadamente grasa, un elevado apetito, altos niveles de leptina circulante, un alto potencial lipogénico y un crecimiento lento. Por otro lado, la raza Landrace presenta un fenotipo opuesto en lo que a estos caracteres se refiere. Es una raza magra, con poco contenido graso, de cuerpo alargado y de rápido crecimiento. Ambas razas porcinas presentan fenotipos extremadamente divergentes para los caracteres de interés, lo que hace que el cruce de ambas líneas genere un material muy valioso para el estudio de estos caracteres complejos (Serra et al. 1998). A partir de la población inicial se han desarrollado generaciones F1, F2, F3 y dos retrocruces (RC1 y RC2) (Figura 7). El RC1 se generó a partir del cruce de cuatro verracos F2 con 22 hembras Landrace y el RC2 mediante la inseminación de 25 hembras Landrace con semen congelado de cinco verracos F1.

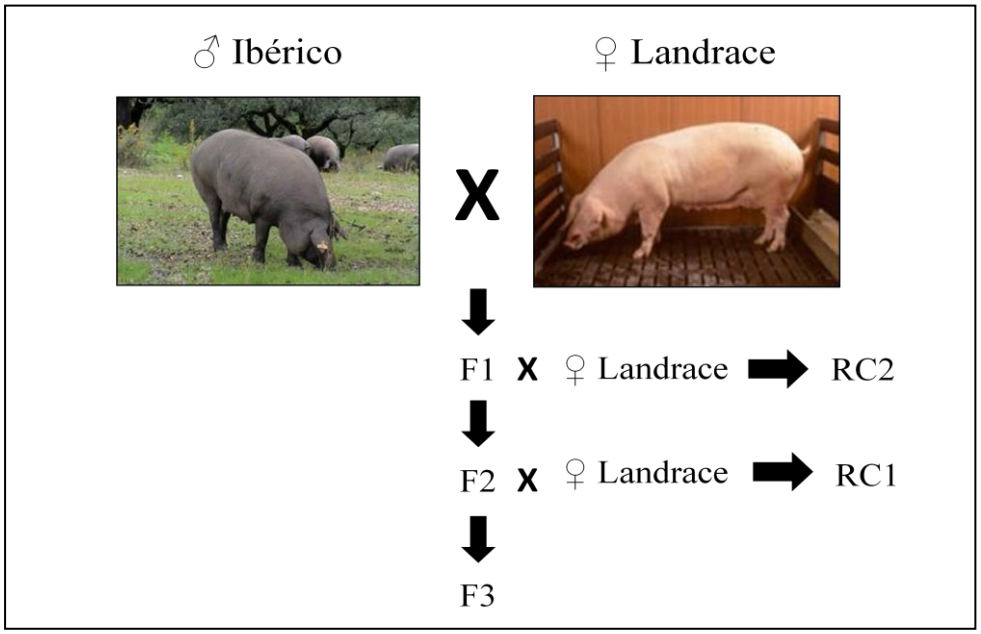


Figura 7. Representación esquemática de los cruces de la población IBMAP.

Se tomaron muestras de ADN total de toda la población y registros fenotípicos de un gran número de caracteres relacionados con el crecimiento y la deposición grasa. Además se tomaron muestras de diversos tejidos de los animales del retrocruce 2 (RC2), lo que ha permitido llevar a cabo estudios de expresión génica. Los estudios previos llevados a cabo en este material han permitido la identificación de un gran número de genes y regiones del genoma asociadas a caracteres relacionados con el crecimiento, la composición corporal, el contenido de ácidos grasos y la calidad de la carne (Ovilo et al. 2002 y 2005, Clop et al. 2003, Muñoz et al. 2007, 2009, 2012 y 2013, Estellé et al. 2006 y 2009, Corominas et al. 2013). Entre los resultados más relevantes cabe destacar los QTLs detectados para calidad de carne localizados en los cromosomas SSC3, SSC4, SSC6, SSC7 y SSC8 y para composición de ácidos grasos en grasa dorsal e intramuscular en el SSC4, SSC8, SSC10 y SSC12 (Ovilo et al. 2002, Clop et al. 2003, Muñoz et al. 2013). Además se han realizado estudios de asociación en diversos genes candidatos mostrando interesantes resultados, como la asociación del polimorfismo *FABP5*g.3000T>G del gen *FABP5* con la deposición grasa (Estellé et al. 2006), la del SNP *ACACA*c.5634T>C del gen *ACACA* con la composición de ácidos grasos en grasa intramuscular (Muñoz et al. 2007, 2009), la mutación no sinónima p.Phe840Leu del gen *MTTP* asociada también con el perfil de ácidos grasos en grasa intramuscular (Estellé et al. 2009), la asociación del SNP ASGA0054039 del gen *PCTP* con el porcentaje de ácido palmítico en grasa dorsal (Muñoz et al. 2012) o la asociación del polimorfismo *ELOVL6*c.533C>T del gen *ELOVL6* con el porcentaje de ácido palmítico y palmitoleico en grasa dorsal e intramuscular (Corominas et al. 2013).

Entre los resultados más relevantes obtenidos en el cruce IBMAP y relacionados con el crecimiento y la deposición grasa cabe destacar la detección de QTLs en los cromosomas SSC2, asociado al área del lomo y al espesor del tocino dorsal, en el SSC4 asociado a la longitud de la canal, al espesor de tocino dorsal y al peso del lomo, las paletas y la panceta, y un tercero en el SSC6 (Figura 8) asociado al área del lomo, al espesor del tocino dorsal y al peso del lomo, las paletas y la panceta, (Óvilo et al. 2000, Varona et al. 2002).

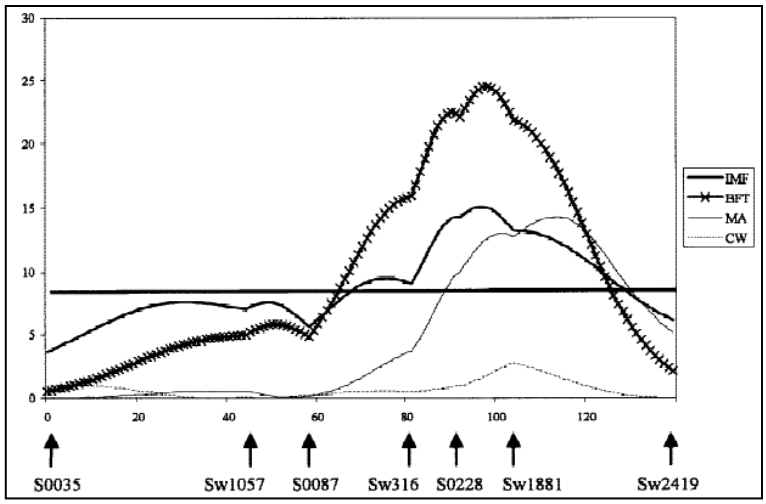


Figura 8. Perfiles de los diferentes test estadísticos (F value) de los cuatro caracteres analizados a lo largo del SSC6 por Óvilo et al. (2000).

El QTL localizado en el SSC6 asociado a deposición grasa y composición corporal ha sido estudiado en mayor profundidad aumentando tanto el número de animales y generaciones, como de marcadores analizados, con el objetivo de refinar esta región QTL. Además, se ha analizado el gen *LEPR* como candidato biológico y posicional a incluir la mutación causal de los efectos de este QTL (Óvilo et al. 2002 y 2005). Este estudio permitió identificar el polimorfismo *LEPR*c.1987C>T fuertemente asociado a caracteres de deposición grasa y crecimiento tanto en el cruce IBMAP (Óvilo et al. 2005) como en otro cruce experimental Ibérico x Meishan (Muñoz et al. 2009), así como en la raza Duroc (Uemoto et al. 2012) y en otros cruces comerciales Duroc x (Landrace x Large White) (Galve et al. 2012) o Duroc x Ibérico (Muñoz et al. 2010), detectándose en todos los casos un fuerte efecto aditivo del alelo T sobre los caracteres asociados. Se ha determinado también que este SNP afecta al nivel de expresión del gen *LEPR* en hipotálamo (Óvilo et al. 2010). Como se ha mencionado anteriormente, el hipotálamo es el principal tejido donde se expresa la isoforma activa del gen *LEPR* desencadenando la cascada de señalización específica de factores de transcripción JAK (Janus kinase) y STAT (Signal transducer and activator of transcription) en respuesta a los niveles de leptina (Figura 9), clave en la regulación del apetito y el balance energético.

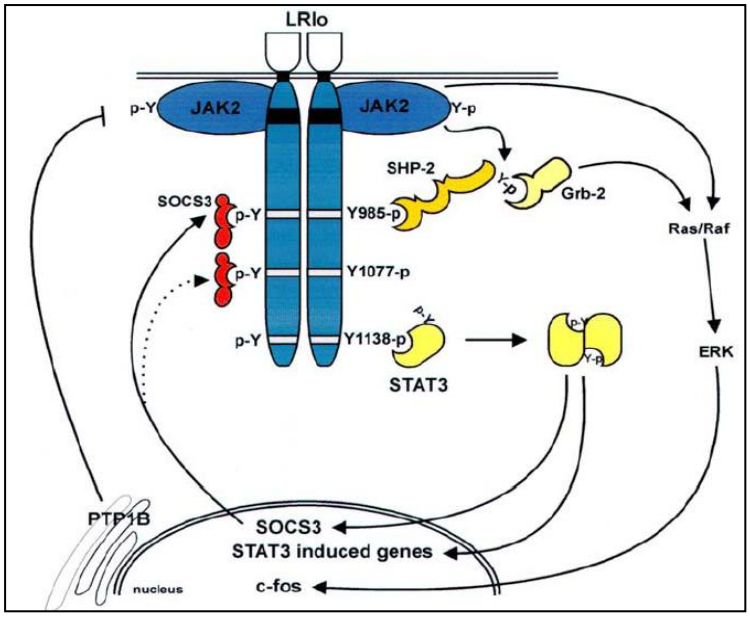


Figura 9. Cascada de señalización activada por la isoforma larga del *LEPR* en el hipotálamo en respuesta a los niveles de leptina (Zabeau et al. 2003).

El alelo T del polimorfismo *LEPR*c.1987C>T está asociado a una reducción de la expresión hipotalámica de este gen probablemente a través de un cambio en la estructura secundaria del ARN mensajero que afectaría a su estabilidad (Óvilo et al. 2010). Además, se detectaron algunos efectos, condicionados por la raza, del genotipo de este SNP sobre la expresión de los genes *NPY* y *CART* que codifican para neuropéptidos involucrados en el control del apetito como se ha descrito anteriormente (Óvilo et al. 2010).

Además de los resultados mencionados en relación al polimorfismo *LEPR*c.1987C>T, se han detectado efectos significativos de otros polimorfismos de este gen sobre el espesor del tocino dorsal en Landrace y Yorkshire, sobre la eficiencia alimentaria en Duroc (Chen et al. 2004), sobre el contenido de grasa intramuscular en un cruce de cerdos coreanos con Yorkshire (Li et al. 2010) y sobre el espesor del tocino dorsal y el porcentaje de carne magra en un cruce Large White x Landrace (Kovácik et al. 2011).

La leptina (*LEP*), ligando del *LEPR*, ha sido también previamente estudiada en porcino debido al importante papel de esta hormona en la señalización relacionada con los procesos descritos de regulación de la ingesta de alimentos y el gasto

energético (Barb et al. 2001) (Figura 10). Pese a su relevancia funcional, los estudios previos de asociación de diversos polimorfismos de este gen con caracteres productivos han mostrado resultados contradictorios (Jian y Gibson 1999, Kennes et al. 2001, Chen et al. 2004, De Oliveira Peixoto et al. 2006, Amils et al. 2008 y Switonski et al. 2010).

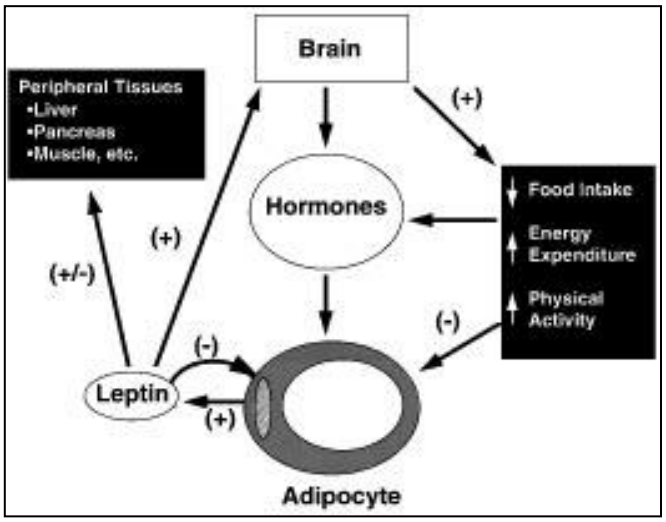


Figura 10. Representación esquemática de la acción de la leptina (Houseknecht et al. 1998)

A partir de estos resultados previos relacionados con el crecimiento y la deposición grasa se ha desarrollado la presente tesis doctoral. Las nuevas generaciones y retrocruces de la población experimental IBMAP, no previamente analizadas, constituyen un excelente material que, junto con las posibilidades que ofrecen las nuevas estrategias de genotipado masivo y análisis global de la expresión génica permiten nuevos enfoques para profundizar en el conocimiento de la base genética de estos caracteres.

OBJETIVOS

El objetivo general de la presente tesis doctoral ha sido profundizar en el conocimiento de la base genética de la regulación del crecimiento y la deposición grasa, caracteres muy importantes desde el punto de vista económico y productivo, pero también por su relación con el desarrollo de patologías humanas relacionadas con la obesidad, para las que la especie porcina es frecuentemente utilizada como modelo. Este objetivo general ha sido abordado desde distintas perspectivas utilizando metodologías tanto de análisis masivo como más tradicionales.

Para llevar a cabo un análisis de detección de QTLs mediante ligamiento se ha explotado la información del chip de genotipado masivo de SNPs porcino (PorcineSNP60BeadChip) de animales de la generación F3 y de los dos retrocruces de la población experimental IBMAP, con el fin de identificar regiones genómicas y genes responsables de la regulación del crecimiento y la deposición grasa en porcino.

> Artículo I: **Análisis de detección de QTLs de crecimiento y composición corporal mediante análisis de ligamiento empleando el chip porcino PorcineSNP60BeadChip.**

Con el objetivo de identificar genes candidato y rutas génicas involucradas en la regulación del crecimiento y la deposición grasa, se ha realizado un análisis del transcriptoma hipotalámico porcino mediante la tecnología de RNA-Seq en animales de uno de los retrocruces de la población IBMAP divergentes para caracteres indicativos de crecimiento y deposición grasa.

> Artículo II: **Identificación de genes reguladores del crecimiento y la deposición grasa en porcino mediante el análisis del transcriptoma hipotalámico.**

Dada la relevancia del gen del receptor de la leptina (*LEPR*) y de su ligando (*LEP*) en la regulación de la ingesta de alimentos, el gasto energético y la composición corporal, y los resultados previos sobre el polimorfismo *LEPR*c.1987C>T, se han realizado estudios estructurales y de expresión génica de ambos genes en el material IBMAP. El objetivo de estos análisis ha sido la identificación de polimorfismos en el gen *LEP* asociados al crecimiento y deposición grasa, analizando su interacción con el polimorfismo LEPRc.1987C>T, y la caracterización de sus patrones de expresión y regulación transcripcional en distintos tejidos. Además se ha analizado la variabilidad de ambos genes en diversas razas porcinas con el fin de profundizar en el conocimiento de la regulación y funciones del eje *LEP*-*LEPR*.

Artículo III: **Estudio del efecto conjunto de polimorfismos de los genes de la leptina y su receptor sobre caracteres productivos.**

Artículo IV: **Diversidad haplotípica de los genes porcinos *LEP* y *LEPR* relacionados con la regulación del crecimiento y la deposición grasa.**

Artículo V: **Caracterización transcripcional de los genes porcinos *LEP* y *LEPR*.**

ARTÍCULO I

Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip

Fernández *et al. BMC Genetics* 2012, **13**:41
http://www.biomedcentral.com/1471-2156/13/41

BMC Genetics

RESEARCH ARTICLE Open Access

Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip

Ana I Fernández[1*], Dafne Pérez-Montarelo[1], Carmen Barragán[1], Yuliaxis Ramayo-Caldas[2], Noelia Ibáñez-Escriche[4], Anna Castelló[2], Jose Luis Noguera[4], Luis Silió[1], Josep María Folch[2,3] and M Carmen Rodríguez[1]

Abstract

Background: The traditional strategy to map QTL is to use linkage analysis employing a limited number of markers. These analyses report wide QTL confidence intervals, making very difficult to identify the gene and polymorphisms underlying the QTL effects. The arrival of genome-wide panels of SNPs makes available thousands of markers increasing the information content and therefore the likelihood of detecting and fine mapping QTL regions. The aims of the current study are to confirm previous QTL regions for growth and body composition traits in different generations of an Iberian x Landrace intercross (IBMAP) and especially identify new ones with narrow confidence intervals by employing the PorcineSNP60 BeadChip in linkage analyses.

Results: Three generations (F3, Backcross 1 and Backcross 2) of the IBMAP and their related animals were genotyped with PorcineSNP60 BeadChip. A total of 8,417 SNPs equidistantly distributed across autosomes were selected after filtering by quality, position and frequency to perform the QTL scan. The joint and separate analyses of the different IBMAP generations allowed confirming QTL regions previously identified in chromosomes 4 and 6 as well as new ones mainly for backfat thickness in chromosomes 4, 5, 11, 14 and 17 and shoulder weight in chromosomes 1, 2, 9 and 13; and many other to the chromosome-wide signification level. In addition, most of the detected QTLs displayed narrow confidence intervals, making easier the selection of positional candidate genes.

Conclusions: The use of higher density of markers has allowed to confirm results obtained in previous QTL scans carried out with microsatellites. Moreover several new QTL regions have been now identified in regions probably not covered by markers in previous scans, most of these QTLs displayed narrow confidence intervals. Finally, prominent putative biological and positional candidate genes underlying those QTL effects are listed based on recent porcine genome annotation.

Keywords: QTL, PorcineSNP60 Beadchip, Growth, Fatness, Body conformation

Background

Hundreds of QTLs have been identified in porcine species (pigQTL database), but there are still relatively few examples for which the mutations that underlie mapped QTLs have been identified [1-4]. The traditional strategy to map QTLs has been to use linkage analysis employing a limited number of microsatellite markers. These analyses usually mapped the QTLs to large intervals, 20 cM or more, which made it difficult to identify the underlying gene and mutation. The success in the positional cloning of these QTLs in domestic animals has been hampered by the absence of high-resolution linkage maps (several markers per cM) [5]. However, the arrival of genome-wide panels of SNPs makes available thousands of markers per chromosome increasing the information content and therefore the likelihood of detecting and fine map QTL regions [5,6].

The Iberian x Landrace experimental cross (IBMAP) was developed to detect QTLs for several economic traits, including growth, fatness and carcass composition [7]. The whole genome QTL scan, carried out in the F2

* Correspondence: avila@inia.es
[1]Departamento de Mejora Genética Animal, INIA, Ctra. De la Coruña km. 7, Madrid 28040, Spain
Full list of author information is available at the end of the article

population using 92 microsatellite markers covering the 18 autosomes, allowed to detect three significant QTLs in SSC2, SSC4 and SSC6 [8]. Subsequent studies and even obtaining new IBMAP generations [9] have delved into knowledge of these regions. Several candidate genes, such as *LEPR, MTTP* and *FABP5,* have been analyzed reporting some successful results [10-16].

Various studies have shown the utility of high-density SNP panels for linkage analyses by providing a greater information content in comparison to microsatellites [6,17-19]. In the present study, we employed the porcine high density SNP panel, PorcineSNP60 BeadChip (Illumina), to carry out a genome QTL scan based on linkage mapping analyses using three of the generations of the IBMAP experimental population. The objective is to confirm previous QTL regions and especially identify new ones with narrow confidence intervals.

Methods

Animals and Phenotypic records

The animals and phenotypic information used in the current study belong to a F3 generation and two different backcrosses of the IBMAP experimental population [9,12]. The IBMAP F1 generation was obtained from three Iberian Guadyerbas boars and 30 Landrace sows. Six F1 boars and 73 F1sows were parents of 577 F2 pigs. Five F1 boars were mated with 25 Landrace sows obtaining 160 backcrossed animals (BC1). In addition three of the F2 boars were mated with 15 F2 sows obtaining 68 animals of the F3 generation, and finally other four F2 boars were mated with 22 Landrace sows obtaining 79 backcrossed animals (BC2). Phenotypic records used in the analyses (Table 1) included the body weight (BW) measured at 150 days of mean age (BW150), and two backfat thickness measures, one at the level of the fourth rib at 4 cm of the midline using ultrasounds at 75 kg of mean weight (BFT75) and the other taken with a rule at slaughter (BFTS). Additionally, intramuscular fat content (IMF) measured by NIRS in *longissimus dorsi* samples and weights of primary cuts (hams, HW, shoulders, SW and loin bone-in, LBW) were also registered at slaughter.

All animal procedures were carried out according to Spanish and European animal experimentation ethics law and approved by the institutional animal ethics committee of IRTA.

SNP data

The 86 F3, 79 BC1 and 160 BC2 pigs, and their related animals from F2, F1 and F0 generations, 416 pigs in total, were genotyped with the PorcineSNP60 BeadChip (Illumina, Inc.), designed by Ramos et al. [20], using the Infinium HD Assay Ultra protocol (Illumina, Inc.). GenomeStudio software (Illumina, Inc.) was employed for visualize, edit and filter the genotyping data. Raw individual data had high-genotyping quality (call rate >0.99). The SNPs filtering was carried in our previous study [21]. Briefly, those SNPs with GenTrain Score lower than 0.85, non-Mendelian inheritance, minor allele frequency less than 0.15, located in sex chromosomes, unmapped in the Sscrofa10 assembly or showing position errors in the linkage mapping were discarded using Plink software [22]. A total of 28,633 SNPs were retained in the dataset after quality control and filtering. In addition, a selection of the most informative SNPs was carried out based in their genetic distance according to the linkage maps generated in our previous study [21]. When the genetic distance among contiguous SNPs was 0, one of them was retained as representative of the linkage group for further analyses.

QTL scan

The linkage maps used for the QTL scan were obtained in Muñoz et al. [21] previous study. A joint QTL scan was performed in all BC1, BC2 and F3 animals. Moreover, two separate analyses were carried out in the BC1 and in the F3 + BC2 animals in agreement with their different parental boar origin. The QTL scans were performed with the following basic model:

$$y_{ijk} = S_i + B_j + u_k + bx_k + P_{ak}a + e_{ijk}$$

Table 1 Phenotypic traits recorded from the BC1 (F1 x Landrace), BC2 (F2 x Landrace) and F3 generations of the Iberian x Landrace cross

Description	Trait	BC1generation			F3 + BC2 generations		
		N	Mean	SD	N	Mean	SD
Weight at 150 days (kg)	W150d	160	79.13	10.49	161	81.69	12.64
Backfat thickness at 75 kg (mm)	BFT75	160	12.69	1.50	134	13.35	2.57
Backfat thickness at slaughter (cm)	BFTS	127	2.50	0.69	148	2.27	0.48
Intramuscular fat percentage (%)	IMF	124	2.06	0.70	147	1.08	0.56
Mean weight of hams (kg)	HW	155	10.22	1.39	148	11.44	1.68
Mean weight of shoulders (kg)	SW	155	5.43	0.80	148	4.70	0.72
Weight of bone-in loins (kg)	BLW	154	7.09	1.03	148	7.18	0.93

where y_{ijk} is the ijk^{th} observation for the analyzed trait, S_i and B_j are the systematic effects for sex (male or female) and batch (eight levels in the whole analysis, three or five levels in the analysis of BC1 or F3 + BC2 pigs), u_k is the random polygenic effect of the k^{th} individual, x_k is a covariable (individual age, body or carcass weight in different analyses) and b its respective slope, a is the QTL additive effect; P_{ak} is the additive coefficient calculated as $P_{ak} = Pr(QQ) - Pr(qq)$, the probability of the k^{th} individual being homozygous for alleles of Iberian origin minus the probability of being homozygous of alleles of Landrace origin and e_{ijk} is the random residual. The infinitesimal genetic effect was treated as random, with covariance $A\sigma_u^2$, A being the numerator relationship matrix. A single residual variance is assumed for all generations (F3, BC1 and BC2). A similar model fitting different QTL effects was used for performing complementary analyses to test the hypothesis of two QTLs mapping in different positions of the same chromosome and with effects a_1 and a_2 on the same trait:

$$y_{ijk} = S_i + B_j + u_k + bx_k + P_{a1k}a_1 + P_{a2k}a_2 + e_{ijk}$$

Finally, joint analyses for two traits were performed to test possible pleiotropic effects of some QTL. The used model was equivalent to the basic, but here the (co)variances of the infinitesimal genetic effects are $A \otimes \begin{pmatrix} \sigma^2_{uy} & \sigma_{uyuz} \\ \sigma_{uzuy} & \sigma^2_{uz} \end{pmatrix}$, where $\otimes$ denotes the Kronecker product and the subindices y and z correspond to the traits.

Likelihood ratio tests (LRT) were calculated comparing the full model and a reduced model without the corresponding QTL effect. The nominal P-values were calculated assuming a χ^2 distribution of the LRT with the degrees of freedom given by the difference between the number of estimated parameters in the reduced and full models. Taking the nominal P-values resulting from the simultaneous testing, their q-values were inferred using QVALUE software [23]. The cut-off of significant QTL at the genome and chromosome level was set at q-value < 0.10. The confidence intervals (CI) were calculated at 95 % following Mangin et al. [24].

Gene annotation

The physical positions of the SNPs were conducted following Sscrofa10.2 genome annotation. The SNP framing the QTL confidence intervals were used to explore gene contain in pig genome assembly 10.2. Gene annotations were retrieved from Gbrowse (www.animalgenome.org/cgi-bin/gbrowse/pig10/).

Association analyses

Complementary association analyses were performed for specific SNPs (and haplotypes) mapped within candidate genes and included in the porcine chip. Candidate genes were identified based on their position within the QTL intervals and their functional relation with the analyzed traits. By the comparison of the SNP position with the gene position, both following Sscrofa10.2, SNPs within the candidate gene were identified and association analyses were conducted. Haplotypes were determined using Haploview software [25].

The analyses were carried out using the standard animal model:

$$y_{ijk} = S_i + B_j + u_k + bx_k + \lambda_k g + e_{ijk}$$

where λ_k is the vector that includes an indicative variable related with the number of copies of one of the SNP or haplotype alleles, which takes 1 or -1values when the k^{th} animal was homozygous for each allele or 0 if the animal was heterozygous; g represented the additive effect of the SNP or haplotype.

All the statistical analyses were performed using the Qxpak v.5.1 software [26].

Results

A total of 8,417 SNPs evenly spaced were used for the analyses. The mean distance between SNPs ranged from 0.18 cM in SSC11 to 0.33 cM in SSC6 (Table 2). The QTL scan has allowed to confirm QTL regions previously identified in the IBMAP population as well as identify new ones (Table 3) and many others at chromosome-wide significant level that are considered as suggestive (Additional file 1: Table S1).

The joint scan of both populations (BC1, F3 + BC2) revealed QTL regions in ten of the 18 autosomes (Table 3 and Additional file 1: Table S1). Six of which were significant to the genome-wide level: three QTLs for BFT75 in SSC4, SSC11 and SSC17, two for SW in SSC1 and SSC4 and one for BLW in SSC4 (Table 3). A complementary analysis was carried out in order to test possible pleiotropic effects of the SSC4 QTL for SW and BLW (Table 3). The results showed a significant pleitropic QTL (P-value = 5.7 x 10^{-6}) at 60 cM with additive effects on these traits (-0.24 ± 0.05 kg and -0.28 ± 0.07 kg, respectively).

The QTL detection analyses carried out in the BC1 generation revealed QTL regions in 14 of the 18 porcine autosomes (Tables 3 and Additional file 1: Table S1), four of which were significant to the genome-threshold level. These genome–wide QTLs were identified in SSC4, SSC11, SSC14 and SSC17 for BFT75 trait (Table 3).

The QTL scan in the F3 + BC2 generations showed QTL regions in 11 of the 18 porcine autosomes (Tables 3 and Additional file 1: Table S1). Ten of these were significant to the genome-wide level: for BFTS in SSC4, SSC5, SSC6 and SSC14, for SW in SSC2, SSC4, SSC6, SSC9 and

Table 2 Molecular information used for the QTL scan

SSC	Number of SNPs	Physical length Mb	Genetic length cM	Mean distance cM between SNPs
		(Mb)	(cM)	
1	490	288.85	139.72	0.29
2	490	156.43	118.58	0.24
3	490	135.91	118.02	0.24
4	490	138.50	119.70	0.24
5	490	108.35	115.81	0.24
6	457	156.97	149.98	0.33
7	490	131.86	125.94	0.26
8	490	147.14	118.64	0.24
9	490	151.45	141.92	0.29
10	410	75.84	111.91	0.27
11	490	82.42	89.59	0.18
12	374	63.85	91.46	0.24
13	490	210.60	106.10	0.22
14	490	153.45	112.15	0.23
15	490	147.44	116.31	0.24
16	490	84.87	82.88	0.17
17	454	68.14	90.52	0.20
18	352	59.26	71.53	0.20

Table 3 Positions, confidence intervals and additive effects of detected significant QTL at the genome-wide level (q-value < 0.10)

Trait	SSC	Position cM (CI)	a (SE)	*P*-value
Whole-population				
BFT75	4	104 (102-109)	0.901 (0.20)	8.4×10^{-6}
	11	26 (25-27)	−0.657 (0.19)	5.7×10^{-4}
	17	58 (56-60)	0.602 (0.19)	1.0×10^{-3}
SW	1	104 (102-104)	−0.221 (0.07)	1.0×10^{-3}
	4	60 (57-62)	−0.316 (0.07)	9.2×10^{-6}
BLW	4	53 (50-54)	−0.300 (0.07)	6.0×10^{-5}
BC1 generation				
BFT75	4	104 (102-109)	1.16 (0.23)	1.0×10^{-6}
	11	26 (25-27)	−0.86 (0.22)	1.8×10^{-4}
	14	112.5 (111-113)	0.70 (0.22)	2.0×10^{-3}
	17	59 (56-61)	0.74 (0.23)	1.9×10^{-3}
F3 + BC2 generations				
BFTS	4	82 (72-91)	0.24 (0.08)	3.0×10^{-3}
	5	73 (71-88)	0.31 (0.09)	9.6×10^{-4}
	6	123 (121-125)	0.32 (0.09)	3.9×10^{-4}
	14	110 (109-113)	0.38 (0.11)	7.4×10^{-4}
SW	2	114 (112-116)	0.45 (0.11)	2.8×10^{-4}
	4	61 (60-64)	−0.35 (0.08)	1.0×10^{-5}
	6	135 (134-136)	−0.34 (0.11)	2.0×10^{-3}
	9	111 (106-113)	−0.23 (0.08)	2.7×10^{-3}
	13	49 (47-53)	−0.28 (0.09)	3.0×10^{-3}
BLW	2	116 (114-117)	0.74 (0.16)	1.3×10^{-5}

SSC13 and for BLW in SSC2 (Table 3). A complementary analysis was carried out in order to test pleiotropic effects of the SSC2 QTL for SW and BLW (Table 3). The results showed a significant pleitropic QTL (*P*-value = 1.1×10^{-4}) at 116 cM with additive effects on these traits (0.40 ± 0.11 kg and 0.74 ± 0.16 kg, respectively).

The separate analyses, using the same set of SNPs markers, evidenced differences between the populations. Examples of these differences are shown in Figure 1. The joint analyses allowed to capture some of the QTLs identified in the separates analysis (in SSC4, SSC11 and SSC17 for BFT75). Nevertheless, a QTL in SSC1 for SW reached the genome-wide significance in the joint analysis but not in the separate ones (Figure 2).

Discussion

The genome wide association study (GWAS) is the approach widely used for the analysis of high density SNP data. In the present study a classical QTL scan, based on the parent line origin assuming alternative alleles fixed in each of the parental populations, has been considered appropriate for the QTL detection analysis in agreement with the experimental design. The QTL scan using this high density panel of 8,417 SNPs has allowed the confirmation of QTL regions previously identified in the IBMAP population. Moreover, new QTLs have been detected, despite using a limited number of animal data, in regions probably not covered by the limited number of microsatellite markers used in previous studies.

Two different QTL analyses were carried out, a joint QTL scan and two separate analyses in agreement with the different parental boar origin of the generations. The separate analyses evidenced the differences between populations regarding the expected QTL genotypes and the random sampling of the QTL alleles in F1 and F2 boars. While only Qq and qq genotypes, coming from the F1 boars, are expected for the QTLs in the BC1 animals, the three possible QTL genotypes (QQ, Qq and qq, coming from the F2 boars) are possible in BC2 and F3 pigs. In addition, the F2 boars used for F3 and BC2 were selected conditioned on their potential genotypes for different QTL (mainly the QTL for growth and fatness in SSC4 and SSC6); however no selection could be done for the F1 boars used for BC1. These differences between populations are reflected in the results obtained. The joint analyses allowed to capture some of the QTLs identified in the separates analysis (in SSC4, SSC11 and SSC17 for BFT75) but not most of them. Nevertheless, other QTL, the one detected in SSC1 for

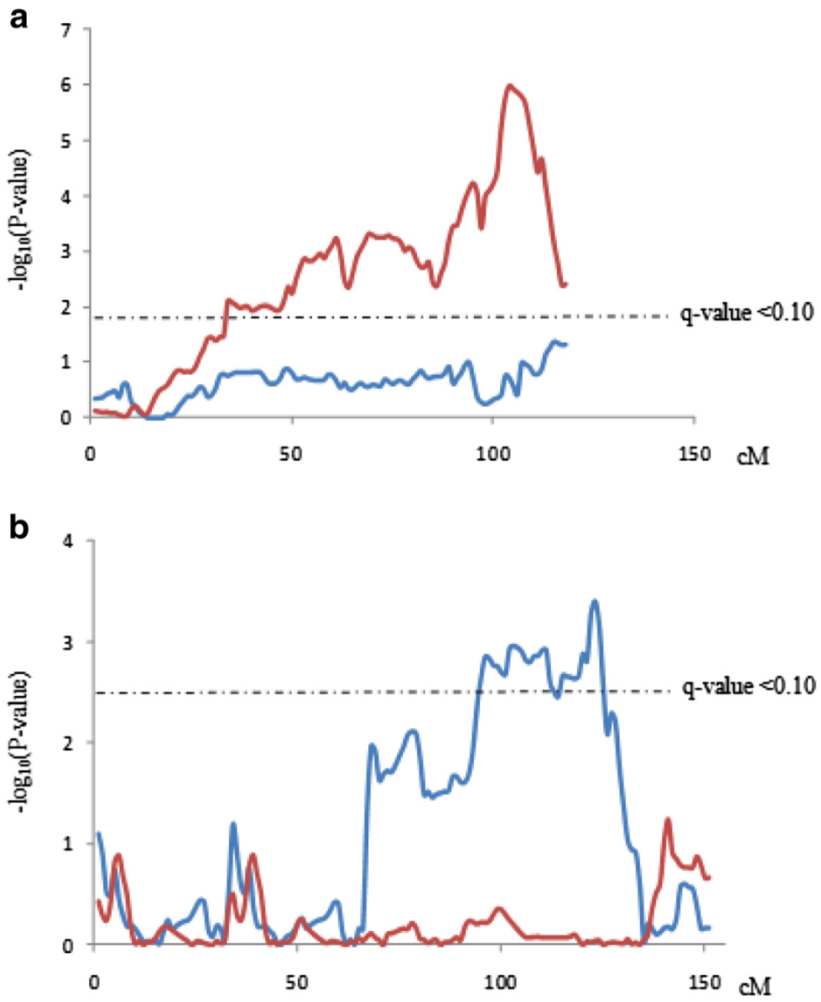


Figure 1 QTL significant profiles in the separate population analyses of SSC4 for BFT75 (a) and SSC6 for BFTS (b). Red lines represent the QTL significant profiles in the BC1 generation and blue lines in the F3 + BC2 generations.

SW, reached the genome-wide significance in the joint analysis but not in the separate one, indicating a gain of detection power with the increase of the record number for this QTL.

The most significant QTL region identified in the present study corresponded to the detected in SSC4 for BFT75, SW and BLW in the joint analysis. The likelihood profiles, shown as $-\log_{10}(P\text{-value})$, showed the

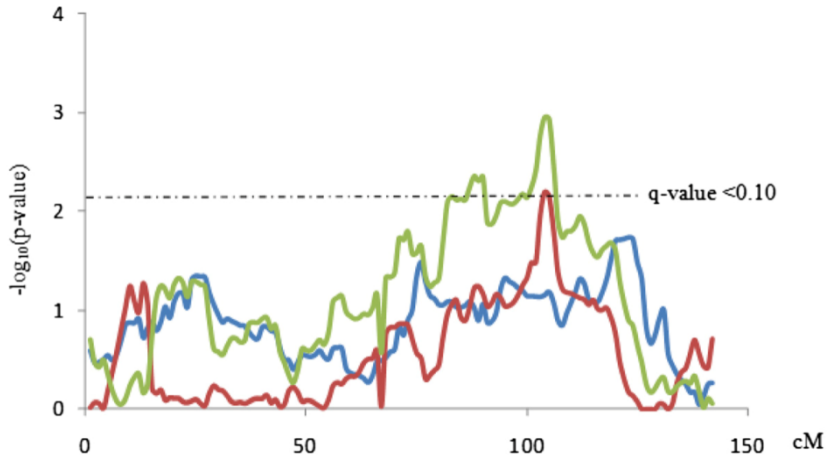


Figure 2 QTL significant profiles in SSC1 for SW in the separate F3 + BC2 and BC1 and in the whole population. Green line represents the QTL significant profile in the whole population, red line in the BC1 generation and blue line in the F3 + BC2 generations.

presence of at least two QTL regions in SSC4, with a maximum located around 53-60 cM position for SW and BLW and around 104 cM for BFT75 (Figure 3). The first QTL region as well as the effects agree with one of the most relevant QTL for growth and fatness described in the IBMAP material for growth and fatness [8,9], overlapping the known *FAT1* region [27,28]. Moreover, a complementary analysis revealed that this first QTL region presented pleotropic effects on SW, BLW and BFTS. However, the second QTL region, around 104 cM affecting BFT75 has been identified for the first time in the IBMAP material, but it has been already described in other populations (Table 4). In addition, the QTL significant profile of SSC4 scan for BFT75 in the BC1 may indicate another potential QTL region around 75 cM (Figure 3). Nevertheless, a complementary analysis employing a model with two QTL *vs* one single QTL did not allow to detect this possible secondary QTL. Another of the most relevant QTL regions for growth and fatness previously identified in the IBMAP experimental population was located around *LEPR* gene in SSC6 [10]. In the present analyses, this QTL has also been detected but only in the F3 + BC2 population (Figure 1). The QTL effects agree with the previously described for backfat thickness and shoulder weight. The *Q* Iberian allele led to an increase of the backfat and a decrease of the shoulder weight. This QTL could not be detected in the BC1, probably due to the lack of QTN segregation in this animal material. In fact, a previous study of the putative causal mutation of this QTL, *LEPR* c.1987 C > T, on this BC1 material, could not reveal

Table 4 New QTL regions identified in the present study *vs* QTLdb and GWAS analysis performed by Fan et al. study [29]

SSC	Position cM	Trait QTLdb	Reference
1	102-104	BFT	[29-31]
		BW	[32]
		ABDF	[33]
		ADG	[32,34]
2	112-117	ADG	[32,35]
4	102-109	10RIBBFT	[36]
5	71-88	BFT	[29,32,37-39]
		BYLEAN	[40]
		ADG	[41]
9	106-113	BW	[42]
		BFT	[37,42]
11	25-27	BFT	[38,43]
		BELLYWT	[30]
13	47-53	BFT	[44,45]
		ADG	[43]
14	109-113	BELLYWT	[37]
		BFT	[29]
17	56-61	ADG	[46]
		BW	[46]
		FATCP	[47]
		FEEDCON	[48]
		HW	[48]

BFT: backfat thickness, BW: body weight, ABDF: abdominal fat weight; ADG: average daily gain; 10RIBBFT: Backfat at tenth rib; BYLEAN: Belly meat content; BELLYWT: Belly weight; FATCP: fat-cuts percentage; FEEDCON: Feed conversion ratio; HW: ham weight.

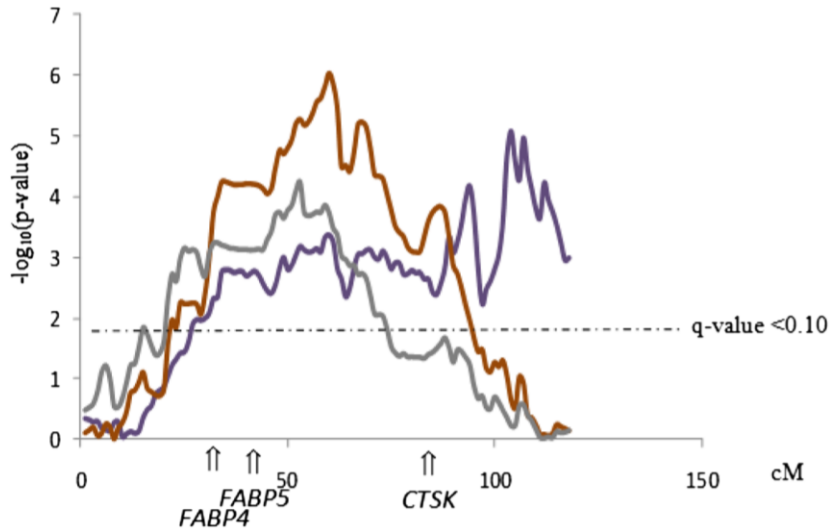


Figure 3 QTL significant profiles in SSC4 for BFT75, SW and BLW traits in the whole-population analysis. Purple line represents the QTL significant profile for BFT75, brown line for SW and gray line for BLW.

significant associations due to the scarcity of some genotypes [13]. The remaining significant genome-wide QTL regions identified in the porcine chromosomes 1, 2, 5, 9, 11, 13, 14 and 17 in one or both populations have been previously associated with growth, fatness and conformation traits in the porcine species (Table 4). However, they were not detected in previous IBMAP scans, except the QTL in SSC9 for SW and in SSC14 for BFT but at suggestive signification level [8]. No genome-wide QTLs could be detected for W150d, IMF and HW traits, probably due to the limited size of the analyzed data, as several ones were detected to chromosome-wide significance level (Additional file 1: Table S1).

Most of the significant QTL regions identified in the present study displayed CI shorter than 5 cM (Table 3), which should facilitate the identification of suitable candidate genes. Thanks to the huge effort of Swine Genome Sequencing Consortium on porcine genome assembly and SNP annotation, a refined search of positional candidate genes could be carried out using Gbrowse tool (Table 5). It should be noted that porcine gene annotation is still scarce, several genes are projected but not annotated; nevertheless gene position within the QTL is more likely to be correct that those inferred from comparative humane-porcine mapping used in previous works. In addition, many annotations discrepancies exist between databases, most likely due to mapping differences between porcine genome versions. In the current study we have mainly used, assuming as more reliable, the latest Sscrofa10.2 genome version for gene annotation. The following paragraphs expose a discussion of the positional and biological putative candidate genes for each of the genome-wide significant QTL regions, according to current porcine gene annotation and available biological information.

The CI (102-104 cM, 232.7-240.4 Mb) of the SSC1 QTL for SW includes 11 protein-coding genes, however only two are annotated to known genes (Table 5). Although no study in porcine has been focused on *PTPRD* gene, studies in human suggest that *PTPRD* gene could play a relevant role in glucose homeostasis and insulin sensitivity [49].

Within the pleiotropic CI (112-117 cM, 150.9-158.3 Mb) of SSC2 QTL for SW and BLW, there are 63 protein-coding genes, 19 out of them are annotated to known genes (Table 5). Among these, *PPARGC1B* constitutes a strong candidate, although it has never been studied in porcine species. PPARGC1B belongs to the PGC-1 family, which act as coactivators in the dysregulation in diseases such as diabetes, obesity and cardiomyopathy in humans [50].

As it was mentioned before, the SSC4 QTL for fatness and conformation traits was identified in previous IBMAP scans [7,8]. In addition, subsequent studies have aimed to deepen the knowledge of this region and some candidate genes (Figure 3) have been analyzed reporting different results [15,16,51-53]. In the present study, apart from that region around 60 cM, another QTL region for BFT75 has been identified around 104 cM position. Within the CI of this second QTL (102-109 cM, 129.7-134.4 Mb) there are 18 protein-coding genes, 12 out of them are annotated to known genes (Table 5), highlighting the *AGL* and *VCAM1* genes as powerful biological candidates underlying the QTL effects. Han et al. [54] study revealed associations of an indel polymorphism in the *AGL* gene with growth, fatness and carcass traits in

Table 5 Annotated genes within the confidence intervals of the new QTLs identified in the present study according to Gbrowse tool

SSC	Position		
	Linkage (cM)	Sscrofa10.2 (Mb)	Official gene symbol
1	102-104	232.67-240.39	*TYRP1*, ***PTPRD***
2	112-117	150.90-158.32	*ARHGAP26, NR3C1, LARS, RBM27, POU4F3, SLC6A7, CSF1R, HMGXB3,* ***PPARGC1B****, PPP2R2B, DPYSL3, JAKMIP2, FBXO38, HTR4, SH3TC2, AFAP1L1, ARHGEF37, CSNK1A1, PDGFR*
4	102-109	129.65-134.38	***VCAM1****, SLC35A3,* ***AGL****, PALMD, DPYD, CNN3, FRRS1, SNX7, PTBP2, ACN9, RNDD3, TMEM56, A4H2R6*
5	71-88	69.45-83.72	*TEAD4, FOXM1, FKBP4, WASH1,SLC6A3, KDM5A, WNK1, RAD52, ERC1,* ***ADIPOR2****, CACNA2D4, ATP6U1E1, BCL2L13, BID, MICAL3, USP18, CPNE8, KIF21A, ABCD2, IRAK4, TMEM117, ANO6, ARID2, SCAF11, SLC38A1, SLC34A4,* ***VDR****, TMEM106C, COL2A1, SENP1, LALBA, APLP2, ALDH1L2, TXNRD1*
9	106-113	120.96-127.05	*NOBOX, TMEM183A,* ***MYOG****, MYBPH, CHI3L1, CHIT1, PRRC2C, MYOC, VAMP4, METTL13, PIGC,* ***FASLG****, TNFRSF4*
11	25-27	26.44-27.89	*KBTBD6, MTRF1, PCDH8*
13	47-53	83.36-91.03	*AMOTL2, ANAPC13, MSL2, PCCB, STAG1, TMEM22,* ***NCK1****, RASA2, GRK7, XRN11, CLDN18, ESYT3, CEP70, FAIM, PIK3CB, FOXL2, COPB2, TRPC1, RBP2, RBP1, CLSTN2, TRIM42, SLC25A36, ACPL2, ZBTB38*
14	109-113	149.48-153.59	*MGMT, GLRX3, PWWP2B, INPP5A, KNDC1, ADAM8, ZNF511,* ***CYP2E1***
17	56-61	39.58-42.08	*TRIB3, TBC1D20, BD129, BD125, REM1,* ***ID1****, BCL2L1, MYLK2, TPX2, TM9SF4, PLAGL2, POFUT1, ASXL1, DNMT3B, MAPRE1, SUN5, BP1F cluster, CDK5RAP1, SNTA1, CBFA2T2*

In bold appears the most prominent putative biological candidates.

an F2 population crossbred Landrace and Jeju (Korea) Black pigs. Recently, Fontanessi et al. [55] study revealed associations of one SNP in *VCAM1* gene with backfat thickness in Italian Large White pigs.

The SSC5 QTL for BFTS showed the largest CI (71-88 cM, 69.5-83.7 Mb), including more than 100 protein-coding genes, 34 of which are annotated to known genes (Table 5). Among the long list of putative candidates, *ADIPOR2* and *VDR* genes highlight as powerful biological candidates, although they have never been studied as candidate gene for fatness in porcine species. The ADIPOR2 mediates the increased AMPK and PPAR-alpha ligand activities, as well as fatty acid oxidation and glucose uptake by adiponectin [56]. Human studies suggest that VDR may function as a determinant of muscle strength, fat mass and body weight [57].

The *LEPR* gene is the most powerful candidate underlying the QTL for fatness and conformation traits mapped in SSC6 in the F3 + BC2 generation. In fact, a highly significant association of a polymorphism located in exon 14, *LEPR* c. 1987 C > T, with growth and fatness has been previously found in several generations of the IBMAP population [11,12]. These effects have been also confirmed in other porcine populations [58-61]. Moreover, functional studies have revealed differences in the LEPR mRNA expression levels in hypothalamus conditional on *LEPR* c.1987 C > T genotype [13] in agreement with the potential causal effect of this QTL on growth and fatness.

The CI (106-113 cM, 121.0-127.1 Mb) of the SSC9 QTL for SW includes 44 protein-coding genes, 13 out of them are annotated to known genes (Table 5). Among the potential list of candidates, the *MYOG* gene plays an essential role in the development and differentiation of muscle. Moreover, studies in porcine species have investigated the associations of *MYOG* polymorphisms with carcass composition and meat quality in pigs evidencing significant associations [62,63]. Also, *FASL* gene has been implicated in skeletal myogenesis [64].

The SSC11 QTL for BFT75 showed one of the shortest CI (25-27 cM, 26.4-27.9 Mb). Within this region only nine protein-coding genes are projected, three of which are annotated to known genes (Table 5), however there is not a feasible candidate as the biological function of these genes have not been elucidated yet.

The CI (109-113 cM, 149.5-153.6 Mb) of SSC14 QTL for live backfat deposition and at slaughter includes 43 protein-coding genes. The *CYP2E1* gene appears among the eight annotated to known genes. The *CYP2E1* has been widely studied in pigs regarding boar taint [65-67], however, its relation to porcine lipid metabolism and fatness has never been explored, even if its key role in obesity and insulin resistance phenotypes has been showed in rodents and humans [68,69].

The SSC13 QTL for SW showed a CI of 7 cM (47-53 cM, 83.4-91.0 Mb). Within this interval 64 projected protein-coding genes are mapped, 25 of which are already annotated to know genes (Table 5). Among them, *NCK1* gene is found as a functional candidate. This gene encodes for a protein implicated in regulating the unfolded protein response, which secondary to obesity impairs glucose homeostasis and insulin actions [70].

Finally, the CI (56-61 cM, 39.6-42.1 Mb) of SSC17 QTL for BFT75 contains 64 coding-protein projected genes, 28 of which are annotated to known genes (Table 5). Among the annotated gene list, the *ID1* gene highlights as biological candidate to underlay the QTL effects. Studies in mice suggest that ID1 is a negative regulator of insulin secretion, playing an essential role in the etiology of glucose intolerance, insulin secretory dysfunction, and β-cell dedifferentiation under conditions of increased lipid supply [71].

Additionally, we noted that some SNPs within five genes (*PTPRD, AGL, VCAM1, VDR*, *FASL* and *CYP2E1*) considered positional and biological candidates, as it is mentioned in the previous paragraphs, are contained in the PorcineSNP60 BeadChip, according to Sscrofa10.2 annotation. Therefore, these SNPs were tested to underlay the corresponding QTL effects in association analyses (Table 6). The main results of the SNP association were for two of the *AGL* SNPs with BFT75 in the BC1 generation. Even more, a haplotype analysis of the two

Table 6 Results of the association analyses of the SNPs that are contained in the PorcineSNP60 BeadChip and mapped within candidate genes underlying the QTL effects in SSC1, 4, 5, 9 and 14

Gene	SNP	Minor alelle frequency	g (SE)	*P*-value
QTL in SSC1 for SW (Whole population)				
PTPRD	ASGA0005690	0.27	−0.01 (0.06)	0.988
	INRA0005932	0	–	–
QTL in SSC4 for BFT75 (BC1 generation)				
AGL	ASGA0022526	0.19	1.15 (0.51)	0.025
	ASGA0022527	0.04	–	–
	ALGA0028692	0.27	0.87 (0.22)	1.8×10^{-4}
VCAM1	DIAS0002972	0.18	0.84 (0.39)	0.033
QTL in SSC5 for BFTS (F3 + BC2 generations)				
VDR	DIAS0001339	0.43	−0.01 (0.07)	0.881
	MARC0076697	0	–	–
QTL in SSC9 for SW (F3 + BC2 generations)				
FASL	DBUN0000737	0.33	−0.03 (0.07)	0.622
	H3GA0028097	0	–	–
QTL in SSC14 for BFT75 (BC1)				
CYP2E1	UMB10000045	0.14	−0.12 (0.39)	0.755

AGL SNPs (haplotypes: ASGA0022526G-ALGA0028692C (H1), ASGA0022526A-ALGA0028692G (H2) and ASGA0022526A-ALGA0028692C (H3)) revealed higher significant effects than the single SNP analyses (*P*-value = 4.6 x 10^{-6}). The H1 haplotype showed the strongest effect (1.02 ± 0.22 mm). These SNPs are located in non-coding regions and they are likely in linkage disequilibrium with the causative mutation underlying the QTL effects in SSC4.

Conclusions

The arrival of the high-density SNP panels makes available high-resolution linkage maps increasing the information content for the successful QTL identification. In the current study, the use of the PorcineSNP60 BeadChip has allowed to detect significant QTL for fatness and yield cuts in ten autosomes (SSC1, SSC2, SSC4, SSC5, SSC6, SSC9, SSC11, SSC13, SSC14 and SSC17). Two of the QTL regions, in SSC4 and SSC6, had been previously identified in the same animal material, however, the remaining ones were not previously detected probably due to the limited number of microsatellite markers employed in those scans. Moreover, most of the significant QTL regions displayed narrow CI making easier the selection of candidate genes. Finally, prominent putative biological and positional candidate genes underlying those QTL effects are listed based on recent porcine genome annotation.

Additional file

Additional file 1: Table S1. Positions and additive effects of significant QTL at the chromosome-wide level (q-value < 0.10).

Abbreviations

QTL: Quantitative trait loci; SNP: Single nucleotide polymorphism; LEPR: Leptin receptor; MTTP: Microsomal triglyceride transfer protein; FABP5: Fatty acid binding protein 5; TYRP1: Tyrosinase-related protein 1; PTPRD: Protein tyrosine phosphatase, receptor type, D; PPARGC1B: Peroxisome proliferator-activated receptor gamma, coactivator 1 beta; AGL: Amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase; VCAM1: Vascular cell adhesion molecule 1; ADIPOR2: Adiponectin receptor 2; VDR: Vitamin D (1,25- dihydroxyvitamin D3) receptor; AMPK: Protein kinase, AMP-activated, beta 1 non-catalytic subunit; MYOG: Myogenin (myogenic factor 4); FASL: Fas ligand (TNF superfamily, member 6); CYP2E1: Cytochrome P450, family 2, subfamily E, polypeptide 1; NCK1: NCK adaptor protein 1; BPIF: BPI fold containing family C; ID1: Inhibitor of DNA binding 1, dominant negative helix-loop-helix protein.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AIF, JMF, JLN, MCR and LS conceived and designed the experiment. MCR and AIF performed the data analysis and drafted the manuscript. JLN, NIE, AIF, MCR, LS and JMF collected the samples and records. AC performed the SNP genotyping, CB, AIF and YR performed the SNP filtering. DPM and AIF performed the SNP probe and gene annotation. All authors read and approved the final manuscript.

Acknowledgements

This work was funded by MICINN projects AGL2008-04818-C03/GAN and CSD2007-00036. DPM was funded by a FPI Ph.D grant from the Spanish Ministerio de Educación (BES-2009-025417). YR was funded by a FPU Ph.D grant from the Spanish Ministerio de Educación (AP2008-01450). We want to thanks to Dr. Martien Groenen (Wageningen, NL) for the SNP annotation on porcine genome assembly, to Anna Mercadé for her technical assistance with the SNPs genotyping and to Rita Benítez and Fabián García for technical support.

Author details

[1]Departamento de Mejora Genética Animal, INIA, Ctra. De la Coruña km. 7, Madrid 28040, Spain. [2]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, UAB, 08193, Bellaterra, Spain. Present address: Consorci CSIC-IRTA-UAB (Centre de Recerca en Agrigenòmica), Edifici CRAG, Campus UAB, Bellaterra, Spain. [3]Centre for Research in Agricultural Genomics (CRAG), Consortium CSIC-IRTA-UAB-UB. Edifici CRAG, Campus Universitat Autonoma Barcelona, 08193, Bellaterra, Spain. [4]Genètica i Millora Animal, IRTA Lleida, 25198, Lleida, Spain.

Received: 27 February 2012 Accepted: 30 April 2012
Published: 20 May 2012

References

1. Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, O'Brien PJ, MacLennan DH: **Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia.** *Science* 199, **253:**448–451.
2. Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundström K, Reinsch N, Gellin J, Kalm E, Roy PL, Chardon P, Andersson L: **A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle.** *Science* 2000, **19:**1248–1251.
3. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L: **A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig.** *Nature* 2003, **425:**832–836.
4. Ciobanu DC, Bastiaansen J, Lonergan SM, Thomsen H, Dekkers JCM, Plastow GS, Rothschild MF: *New alleles in calpastatin gene are associated with meat quality traits in pigs J Anim Sci* 2004, **82:**2829–2839.
5. Andersson L, Georges M: *Domestic-animal genomics: deciphering the genetics of complex traits: Nat Rev Genet* 2004, **5:**202–212.
6. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC: **Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites.** *Am J Hum Genet* 2004, **75:**54–64.
7. Pérez-Enciso M, Varona L: **Quantitative trait loci mapping in F2 crosses between outbred lines.** *Genetics* 2000, **155:**391–405.
8. Varona L, Óvilo C, Clop A, Noguera JL, Pérez-Enciso M, Coll A, Folch JM, Barragán C, Toro MA, Babot D, Sánchez A: **QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects.** *Genet Res* 2002, **80:**145–154.
9. Mercadé A, Estellé J, Noguera JL, Folch JM, Varona L, Silió L, Sánchez A, Pérez-Enciso M: **On growth, fatness, and form: a further look at porcine chromosome 4 in an Iberian x Landrace cross.** *Mamm Genome* 2005, **16:**374–382.
10. Óvilo C, Pérez-Enciso M, Barragan C, Clop A, Rodriguez C, Oliver MA, Toro MA, Noguera JL: **A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6.** *Mamm Genome* 2000, **11:**344–346.
11. Óvilo C, Oliver A, Noguera JL, Clop A, Barragán C, Varona L, Rodríguez C, Toro MA, Sánchez A, Pérez-Enciso M, Silió L: **Test for positional candidate genes for body composition on pig chromosome 6.** *Genet Sel Evol* 2002, **34:**465–479.
12. Óvilo C, Fernández A, Noguera JL, Barragán C, León R, Rodríguez C, Mercadé A, Alves E, Folch JM, Varona L, Toro MA: **Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross.** *Genet Res* 2005, **85:**57–67.
13. Óvilo C, Fernández A, Fernández AI, Folch JM, Varona L, Benítez R, Nuñez Y, Rodríguez C, Silió L: **Hypothalamic expression of porcine leptin receptor**

(LEPR), neuropeptide Y (NPY), and cocaine- and amphetamine-regulated transcript (CART) genes is influenced by LEPR genotype. *Mamm Genome* 2010, **21:**583–591.
14. Estellé J, Fernández AI, Pérez-Enciso M, Fernández A, Rodríguez C, Sánchez A, Noguera JL, Folch JM: **A non-synonymous mutation in a conserved site of the MTTP gene is strongly associated with protein activity and fatty acid profile in pigs.** *Anim Genet* 2009, **40:**813–820.
15. Estellé J, Pérez-Enciso M, Mercadé A, Varona L, Alves E, Sánchez A, Folch JM: **Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross.** *Anim Genet* 2006, **37:**589–591.
16. Mercadé A, Pérez-Enciso M, Varona L, Alves E, Noguera JL, Sánchez A, Folch JM: **Adipocyte fatty-acid binding protein is closely associated to the porcine FAT1 locus on chromosome 4.** *J Anim Sci* 2006, **84:**2907–2913.
17. Evans DM, Cardon LR: **Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps.** *Am J Hum Genet* 2004, **75:**687–692.
18. Gonzalez-Neira A, Rosa-Rosa JM, Osorio A, Gonzalez E, Southey M, Sinilnikova O, Lynch H, Oldenburg RA, van Asperen CJ, Hoogerbrugge N, Pita G, Devilee P, Goldgar D, Benitez J: **Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies.** *BMC Genomics* 2007, **8:**299.
19. Chioza BA, Aicardi J, Aschauer H, Brouwer O, Callenbach P, Covanis A, Dooley JM, Dulac O, Durner M, Eeg-Olofsson O, Feucht M, Friis ML, Guerrini R, Kjeldsen MJ, Nabbout R, Nashef L, Sander T, Sirén A, Wirrell E, McKeigue P, Robinson R, Gardiner RM, Everett KV: **Genome wide high density SNP-based linkage analysis of childhood absence epilepsy identifies a susceptibility locus on chromosome 3p23-p14.** *Epilepsy Res* 2009, **87:**247–255.
20. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TP, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MA: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.** *PLoS One* 2009, **4:**e6524.
21. Muñoz M, Alves E, Ramayo-Caldas Y, Casellas J, Rodríguez C, Folch JM, Silió L, Fernández AI: **Recombination rates across porcine autosomes inferred from high-density linkage maps.** *Anim Genet* 2011, doi:10.1111/j.1365-2052.2011.02301.x.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81:**559–275.
23. Storey JD, Taylor JE, Siegmund D: **Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach.** *J R Stat Soc Series B* 2004, **66:**187–205.
24. Mangin B, Goffinet B, Rebaï A: **Constructing confidence intervals for QTL location.** *Genetics* 1994, **138:**1301–1308.
25. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21:**263–265.
26. Pérez-Enciso M: **Misztal I.** *Qxpak.5: old mixed model solutions for new genomics problems. BMC Bioinformatics* 2011, **12:**202.
27. Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K, Andersson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, *et al*: **Genetic mapping of quantitative trait loci for growth and fatness in pigs.** *Science* 1994, **263:**1771–1774.
28. Marklund L, Nyström PE, Stern S, Andersson-Eklund L, Andersson L: **Confirmed quantitative trait loci for fatness and growth on pig chromosome 4.** *Heredity* 1999, **82:**134–141.
29. Fan B, Onteru SK, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF: **Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs.** *PLoS One* 2011, **6:**e14726.
30. Rohrer GA, Keele JW: **Identification of quantitative trait loci affecting carcass composition in swine: I.** *Fat deposition traits. J Anim Sci* 1998, **76:**2247–2254.
31. Rohrer GA: **Identification of quantitative trait loci affecting birth characters and accumulation of backfat and weight in a Meishan-White Composite resource population.** *J Anim Sci* 2000, **78:**2547–2553.
32. Bidanel JP, Milan D, Iannuccelli N, Amigues Y, Boscher MY, Bourgeois F, Caritez JC, Gruand J, Le Roy P, Lagant H, Quintanilla R, Renard C, Gellin J, Ollivier L, Chevalet C: **Detection of quantitative trait loci for growth and fatness in pigs.** *Genet Sel Evol* 2001, **33:**289–309.
33. Geldermann H, Cepica S, Stratil A, Bartenschlager H, Preuss S: **Genome-wide mapping of quantitative trait loci for fatness, fat cell characteristics and fat metabolism in three porcine F2 crosses.** *Genet Sel Evol* 2010, **42:**31.
34. Quintanilla R, Milan D, Bidanel JP: **A further look at quantitative trait loci affecting growth and fatness in a cross between Meishan and Large White pig populations.** *Genet Sel Evol* 2002, **34:**193–210.
35. Knott SA, Marklund L, Haley CS, Andersson K, Davies W, Ellegren H, Fredholm M, Hansson I, Hoyheim B, Lundström K, Moller M, Andersson L: **Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs.** *Genetics* 1998, **149:**1069–1080.
36. Sławińska A, Siwek M, Knol EF, Roelofs-Prins DT, van Wijk HJ, Dibbits B, Bednarczyk M: **Validation of the QTL on SSC4 for meat and carcass quality traits in a commercial crossbred pig population.** *J Anim Breed Genet* 2009, **126:**43–51.
37. Edwards DB, Ernst CW, Tempelman RJ, Rosa GJ, Raney NE, Hoge MD, Bates RO: **Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: I.** *Growth traits. J Anim Sci* 2008, **86:**241–253.
38. Milan D, Bidanel JP, Iannuccelli N, Riquet J, Amigues Y, Gruand J, Le Roy P, Renard C, Chevalet C: **Detection of quantitative trait loci for carcass composition traits in pigs.** *Genet Sel Evol* 2002, **34:**705–728.
39. Guo T, Ren J, Yang K, Ma J, Zhang Z, Huang L: **Quantitative trait loci for fatty acid composition in longissimus dorsi and abdominal fat: results from a White Duroc x Erhualian intercross F2 population.** *Anim Genet* 2009, **40:**185–191.
40. Liu G, Kim JJ, Jonas E, Wimmers K, Ponsuksili S, Murani E, Phatsara C, Tholen E, Juengst H, Tesfaye D, Chen JL, Schellander K: **Combined line-cross and half-sib QTL analysis in Duroc-Pietrain population.** *Mamm Genome* 2008, **19:**429–438.
41. de Koning DJ, Harlizius B, Rattink AP, Groenen MA, Brascamp EW, van Arendonk JA: **Detection and characterization of quantitative trait loci for meat quality traits in pigs.** *J Anim Sci* 2001, **79:**2812–2819.
42. Xue HL, Zhou ZX: **Effects of the MyoG gene on the partial growth traits in pigs.** *Yi Chuan Xue Bao* 2006, **33:**992–997.
43. Kim CW, Hong YH, Yun SI, Lee SR, Kim YH, Kim MS, Chung KH, Jung WY, Kwon EJ, Hwang SS, Park DH, Cho KK, Lee JG, Kim BW, Kim JW, Kang YS, Yeo JS, Chang KT: **Use of microsatellite markers to detect quantitative trait loci in Yorkshire pigs.** *J Reprod Dev* 2006, **52:**229–237.
44. Liu G, Jennen DG, Tholen E, Juengst H, Kleinwächter T, Hölker M, Tesfaye D, Un G, Schreinemachers HJ, Murani E, Ponsuksili S, Kim JJ, Schellander K, Wimmers K: **A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population.** *Anim Genet* 2007, **38:**241–252.
45. Xu XL, Xu XW, Pan PW, Li K, Jiang ZH, Yu M, Rothschild MF, Liu B: **Porcine skeletal muscle differentially expressed gene CMYA1: isolation, characterization, mapping, expression and association analysis with carcass traits.** *Anim Genet* 2009, **40:**255–261.
46. Ramos AM, Bastiaansen JW, Plastow GS, Rothschild MF: **Genes located on a SSC17 meat quality QTL region are associated with growth in outbred pig populations.** *Anim Genet* 2009, **40:**774–778.
47. Pierzchala M, Cieslak D, Reiner G, Bartenschlager H, Moser G, Geldermann H: **Linkage and QTL mapping for Sus scrofa chromosome 17.** *J Anim Breed Genet* 2003, **120:**132–137.
48. Russo V, Fontanesi L, Scotti E, Beretti F, Davoli R: **Nanni Costa L, Virgili R, Buttazzoni L: Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs.** *J Anim Sci* 2008, **86:**3300–3314.
49. Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, Wang TY, Chen RH, Shiu CF, Liu YM, Chang CC, Chen P, Chen CH, Fann CS, Chen YT, Wu JY: **A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese.** *PLoS Genet* 2010, **6:**e1000847.
50. Handschin C, Spiegelman BM: **Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism.** *Endocr Rev* 2006, **27:**728–735.
51. Silva KM, Bastiaansen JW, Knol EF, Merks JW, Lopes PS, Guimarães SE, van Arendonk JA: **Meta-analysis of results from quantitative trait loci mapping studies on pig chromosome 4.** *Anim Genet* 2011, **42:**280–292.

52. Fontanesi L, Scotti E, Buttazzoni L: **Dall'Olio S, Davoli R, Russo V: A single nucleotide polymorphism in the porcine cathepsin K (CTSK) gene is associated with back fat thickness and production traits in Italian Duroc pigs.** *Mol Biol Rep* 2010, **37:**491–495.
53. Ojeda A, Estellé J, Folch JM, Pérez-Enciso M: **Nucleotide variability and linkage disequilibrium patterns at the porcine FABP5 gene.** *Anim Genet* 2008, **39:**468–473.
54. Han SH, Shin KY, Lee SS, Ko MS, Jeong DK, Oh HS, Yang BC, Cho IC: **SINE indel polymorphism of AGL gene and association with growth and carcass traits in Landrace x Jeju Black pig F(2) population.** *Mol Biol Rep* 2010, **37:**467–471.
55. Fontanesi L, Galimberti G, Calò DG, Fronza R, Martelli PL, Scotti E, Colombo M, Schiavo G, Casadio R, Buttazzoni L, Russo V: **Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for backfat thickness in Italian Large White pigs using a selective genotyping approach.** *J Anim Sci* 2012, **:** [Epub ahead of print].
56. Yamauchi T, Kamon J, Ito Y, Tsuchida A, Yokomizo T, Kita S, Sugiyama T, Miyagishi M, Hara K, Tsunoda M, Murakami K, Ohteki T, Uchida S, Takekawa S, Waki H, Tsuno NH, Shibata Y, Terauchi Y, Froguel P, Tobe K, Koyasu S, Taira K, Kitamura T, Shimizu T, Nagai R, Kadowaki T: **Cloning of adiponectin receptors that mediate antidiabetic metabolic effects.** *Nature.* 2003, **423:**762–769.
57. Grundberg E, Brändström H, Ribom EL, Ljunggren O, Mallmin H, Kindmark A: **Genetic variation in the human vitamin D receptor is associated with muscle strength, fat mass and body weight in Swedish women.** *Eur J Endocrinol* 2004, **150:**323–328.
58. Lopez-Buesa P, Óvilo C, Rodríguez MC, Varona L, Silió L, Burgos C, Galve A: **The effects of leptin receptor (LEPR) and melanocortin-4 receptor (MC4R) polymorphisms in fat content, fat distribution, and fat composition in a Duroc x Landrace/Large White cross.** *Livestock Science* 2011, **:**. doi:10.1016/j.livsci.2012.01.010.
59. Muñoz G, Óvilo C, Silió L, Tomás A, Noguera JL, Rodriguez MC: **Single- and joint population analyses of two experimental pig crosses to confirm quantitative trait loci on Sus scrofa chromosome 6 and leptin receptor effects on fatness and growth traits.** *J Anim Sci* 2009, **87:**459–468.
60. Muñoz G, Alcázar E, Fernández A, Barragán C, Carrasco A, de Pedro E, Silió L, Sánchez JL, Rodríguez MC: **Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc x Iberian crossbred pigs.** *Meat Sci* 2011, **88:**169–173.
61. Rodriguez MC, Fernandez A, Carrasco C, Garcia A, Gomez E, De Mercado E, Lopez MA, Óvilo C, Silio L: **Effect of LEPR c.2002 C>T SNP on feed intake and growth in heavy Duroc X Iberian crossbred pigs.** *In Proceedings of the ninth world congress on genetics applied to livestock production* 2010, **:**1–6. *Leipzig.*
62. Krzęcio E, Koćwin-Podsiadła M, Kurył J, Zybert A, Sieczkowska H, Antosik K: **The effect of interaction between genotype CAST/RsaI (calpastatin) and MYOG/MspI (myogenin) on carcass and meat quality in pigs free of RYR1(T) allele.** *Meat Sci* 2008, **80:**1106–1115.
63. Wyszyńska-Koko J, Pierzchała M, Flisikowski K, Kamyczek M, Rózycki M, Kurył J: **Polymorphisms in coding and regulatory regions of the porcine MYF6 and MYOG genes and expression of the MYF6 gene in m. longissimus dorsi versus productive traits in pigs.** *J Appl Genet* 2006, **47:**131–138.
64. Ikeda T, Kanazawa T, Otsuka S, Ichii O, Hashimoto Y, Kon Y: **Expression of caspase family and muscle- and apoptosis-specific genes during skeletal myogenesis in mouse embryo.** *J Vet Med Sci* 2009, **71:**1161–1168.
65. Squires EJ: **Lundström: Relationship between cytochrome P450IIE1 in liver and levels of skatole and its metabolites in intact male pigs.** *J Anim Sci* 1997, **75:**2506–2511.
66. Babol J, Squires EJ, Lundström K: **Hepatic metabolism of skatole in pigs by cytochrome P4502E1.** *J Anim Sci* 1998, **76:**822–828.
67. Lin Z, Lou Y, Squires EJ: **Functional polymorphism in porcine CYP2E1 gene: Its association with skatole levels.** *J Steroid Biochem Mol Biol* 2006, **99:**231–237.
68. Zong H, Armoni M, Harel C, Karnieli E, Pessin JE: **Cytochrome P450 CYP2E1 knockout mice are protected against high fat diet induced obesity and insulin resistance.** *Am J Physiol Endocrinol Metab* 2011, doi:10.1152/ajpendo.00258.2011.
69. Aubert J, Begriche K, Knockaert L, Robin MA, Fromenty B: **Increased expression of cytochrome P450 2E1 in nonalcoholic fatty liver disease: mechanisms and pathophysiological role.** *Clin Res Hepatol Gastroenterol* 2011, **35:**630–637.
70. Latreille M, Laberge MK, Bourret G, Yamani L, Larose L: **Deletion of Nck1 attenuates hepatic ER stress signaling and improves glucose tolerance and insulin signaling in liver of obese mice.** *Am J Physiol Endocrinol Metab* 2011, **300:**E423–434.
71. Akerfeldt MC, Laybutt DR: **Inhibition of Id1 augments insulin secretion and protects against high-fat diet-induced glucose intolerance.** *Diabetes* 2011, **60:**2506–2514.

doi:10.1186/1471-2156-13-41
Cite this article as: Fernández *et al.*: Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. *BMC Genetics* 2012 **13**:41.

ARTÍCULO II

Identification of genes regulating growth and fatness traits in pig through hypotlamaic transcriptome analysis

IDENTIFICATION OF GENES REGULATING GROWTH AND FATNESS TRAITS IN PIG THROUGH HYPOTHALAMIC TRANSCRIPTOME ANALYSIS

Dafne Pérez-Montarelo[1*], Ole Madsen[2], Estefânia Alves[1], M. Carmen Rodríguez[1], Josep María Folch[3,4], José Luis Noguera[5], Martien A.M. Groenen[2] and Ana I. Fernández[1].

[1]Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain.

[2]Animal Breeding and Genomics Centre, Wageningen University, 6708 WD Wageningen, The Netherlands.

[3] Genètica Animal, Centre de Recerca en Agrigenòmica (CRAG), Campus UAB, 08193 Bellaterra, Spain.

[4]Ciència Animal i dels Aliments, Universitat Autonoma de Barcelona, 08193 Bellaterra, Spain.

[5] Genètica i Millora Animal, Institut de Recerca i Tecnologies Agroalimentaries, 25198 Lleida, Spain.

*Corresponding author: dafne.perez@inia.es. Phone: (0034)913473792

ABSTRACT:

Previous studies on Iberian x Landrace (IBMAP) pig intercrosses have enabled the identification of several QTL regions related to growth and fatness traits; however the genetic variation underlying those QTLs are still unknown. These traits are not only relevant because of their impact on economically important production traits, but also because pig constitutes a widely studied animal model for human obesity and obesity related diseases. The hypothalamus is the main gland regulating growth, food intake and fat accumulation. Therefore, the aim of this work was to identify genes and/or gene transcripts involved in the determination of growth and fatness in pig by a comparison of the whole hypothalamic transcriptome (RNA-seq) in two groups of phenotypically divergent IBMAP pigs. Around 16,000 of the ~25.010 annotated genes were expressed in these hypothalamic samples, with most of them showing intermediate expression levels. Functional analyses supported the key role of the hypothalamus in the regulation of growth, fat accumulation and energy expenditure. More than 250 differentially expressed (DE) genes and novel transcript isoforms were identified between the two groups of pigs. Twenty-one DE genes/transcripts that co-localized in previously identified QTL regions and/or whose biological functions are related to the traits of interest were explored in more detail. Additionally, the transcription factors potentially regulating these genes and the subjacent networks and pathways were also analyzed. This study allows us to propose strong candidate genes for growth and fatness based on expression patterns, genomic location and network interactions.

KEYWORDS: RNA-Seq, porcine, hypothalamus, growth, fatness.

INTRODUCTION:

Growth and fatness are some of the most economically important traits for meat production in livestock. Phenotypic traits such as average daily gain, backfat thickness and carcass composition are closely related to the efficiency of pig production. Both growth and fatness are the outcome of the interactions

between many genetic and environmental factors. Understanding the genetic pathways regulating these traits may provide new tools that will help to modulate animal growth and efficiency. Moreover, the pig is also being used as a biomedical model for studying human obesity, energy metabolism and diabetes (57). Especially the high similarity in metabolic and digestive features as well as other anatomical and physiological characteristics between human and pig makes pig a favorable animal model compared to rodents in particular when translating results to human. Despite extensive research aimed at identifying the molecular basis of obesity and related diseases, the underlying physiogenetic mechanisms have not been clearly identified (57). Furthermore, the use of pigs as a model offers the possibility of mating phenotypically divergent animals creating populations segregating the traits of interest and subsequently intercrossing the offspring (29). Hence, studying the genetic basis of growth and fatness in the pigs is of interest from both an animal production and biomedical point of view.

Even though the regulation of feeding behavior is very complex and different brain areas are involved, the hypothalamus has been shown to be the main feeding regulatory center, responsible for food intake, energy homeostasis and body weight (19). The hypothalamus is crucial for the detection of nutrient levels and, in response to different brain and peripheral signals from the gut and adipose tissue, it modulates food intake and energy expenditure (56). The hypothalamus consists of several interconnected nuclei, including the arcuate nucleus (ARC), the paraventricular nucleus (PVN), the lateral hypothalamic area (LHA), the ventromedial nucleus (VMH) and the dorsomedial nucleus (DMH) (19, 56). The ARC is the key nucleus, strategically positioned to integrate a number of peripheral signals controlling food intake (56). The signal from ARC projects to secondary neurons in the PVN, LHA, VMH and DMH that are organized into a complex network of orexigenic and anorexigenic neuropeptides that responds to meal-satiety signals for

long term regulation of energy homeostasis (19).

Several QTL regions have been identified in pig associated with growth and fatness (Pig QTLdb; 20), but there are few examples where the genetic variation responsible for the QTL effects have been identified (36, 62). We performed high-throughput RNA sequencing (RNA-Seq) of hypothalamic tissue to identify the genes responsible for these QTL effects. The RNA-Seq technology is a powerful approach, which allows a genome wide transcriptome characterization and identification of the full set of transcripts expressed. RNA-seq provides enormously detailed insight in transcript expression and in the discovery of new transcripts and gene polymorphisms in a single assay. In the last years, several studies have employed RNA-Seq to explore the porcine transcriptome of different tissues such as liver (3, 16, 25, 51), gonads (8, 16), endometrium (54), traqueobronchial lymph nodes (37), *longissimus dorsi* muscle and abdominal fat (3). Previous pig hypothalamic transcriptome analyses have been conducted using microarrays (1, 10, 47, 66) or expressed sequence tags (ESTs) (12) focusing on sex or breed characterization. However, no hypothalamic RNA-seq transcriptome studies have been carried out in pig so far. To our knowledge the only hypothalamic RNA-Seq transcriptome study performed in a livestock species has been conducted in cattle and focused on fertility related traits (11).

The Iberian x Landrace (IBMAP) experimental population was specifically developed to identify QTLs for growth, fatness and body composition traits, because of the large phenotypic divergence of both parental breeds in relation to these traits (46). Several studies based on this population have allowed the detection of QTL regions for the traits of interest (9, 63). Significant QTL regions associated with growth and fatness have been identified on porcine chromosomes (SSC) 2, 4, 5, 6, 14 and 17. Moreover, some candidate genes, such as *LEPR, FABP4* or *FABP5* have been analyzed reporting relevant associations of polymorphisms of these genes (LEPRc.1986C>T, FABP4:g.2634_2635insC, FABP4:g.6252C>T and

FABP5;g.3000T>G) with the traits of interest (7, 45). However, the identification and confirmation of the causal mutations underlying those QTLs have not been successfully solved yet and further information is needed to understand the complex genetic basis of these traits.

In the present study, the RNA-Seq technology has been employed to compare the transcriptome of hypothalamic samples of divergent pigs for growth and fatness traits from the IBMAP population, with the aim to identify genes related to these traits.

MATERIALS AND METHODS:

Animal material, RNA isolation and sequencing

Animal manipulations were performed according to the Spanish Policy for Animal Protection RD1201/05, which meets the European Union Directive 86/609 about the protection of animals used in experimentation. The animal material used in the present study is derived from a backcross (BC) generated from the IBMAP experimental population (44). Briefly, three Iberian boars were mated with 30 Landrace sows (F0) to produce 70 F1 animals. To generate the BC, five F1 boars were mated with 25 Landrace sows and 187 BC animals were obtained. All pigs were grown in an experimental farm under standard conditions. The animals were slaughtered at an age of 175 days. Phenotypic traits related to growth, fatness, and body composition were measured in all BC animals (Supplementary Table 1).

A principal component analysis (PCA) of the BC animals was performed according to four indicators for growth and fatness traits: average daily gain (kg/d), backfat thickness at 90 kg (mm), and percentages of C18:2 in backfat and intramuscular fat. Note that the C18:2 content of fat tissues is positively correlated to pig lean growth. Ten male pigs with extreme phenotypes, according to the first principal component, part of the same batch, were selected for this study and arranged into two groups (Figure 1). The five males showing the highest values of growth and fatness indicators were included in the High (H) group and the five males showing the lowest values for these traits were included in the Low (L) group (Figure 1). The mean values of the

four indicators in the H and L groups were 0.92-0.78 kg/d of average daily gain, 16.2-11.6 mm of backfat thickness, 12.6-16.7 % of C18:2 in backfat, and 8.1-11.9 % of C18:2 in intramuscular fat, respectively.

Hypothalamic samples of the 10 selected animals were collected at slaughter, immediately frozen in liquid nitrogen and stored at -80ºC until analyzed. Total RNA was extracted using the *RiboPure TM of High Quality total RNA* kit (Ambion, Austin, TX, USA) following the manufacturer's recommendations and quantified using a *NanoDrop-100* spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The integrity of the RNA was assessed using an *Agilent 2100 Bioanalyzer* device (Agilent technologies, Santa Clara, CA, USA). The RNA integrity value (RIN) of the samples ranged between 7.1 and 8.1. Paired-end libraries with fragments of 300bp were prepared using the TruSeq RNA Sample Prep Kit v2 (Illumina, Inc., San Diego CA, USA) for each sample. Multiplex sequencing of the libraries was performed on a *Illumina Hi-Seq 2000* (Fasteris SA, Plan-les-Ouates, Switzerland) with three samples per lane, according to the manufacturer's instructions at CNAG (Centro Nacional de Análisis Genómico) generating paired-end reads of 75 bp .

Mapping and assembly

Quality of the raw sequencing data was accessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to quality trim the data with default settings and to remove the sequencing adaptors and poly A and T tails (stringency of 6 bp, -s 6) keeping only paired-end reads where both pairs were longer than 40bp. Filtered reads were mapped against the pig reference genome (Sscrofa10.2) (15) using TopHat v.2.0.5 (59) with Bowtie2 (v.2.0.0.7) applying default settings except that reads were first aligned to the ENSEMBL (10.2.68) transcriptome annotation (-G option) and the distance between both pairs was set to 50bp (inner-mean distance) and the standard deviation at 150bp.

Transcripts were assembled and quantified in FPKM (fragments per kilobase of transcript per million

mapped fragments) by Cufflinks v2.0.2 (60). Additionally, the CummeRbund Bioconductor R package (13) was employed to analyze Cuffdiff (see below) output and determine the clustering of the samples according to the expression data, to assess the consistence of the groups that are being compared. Two of the samples, one from each group, were discarded from the study because their clustering deviate largely from the expected clustering, probably due to sampling or RNA processing problems.

Identification of novel isoforms

Novel expressed isoforms were extracted using the Cuffcompare tool from Cufflinks. Cuffcompare was run using the ENSEMBL (10.2.68) transcriptome annotation as a reference for assessing the accuracy of the predicted Cufflinks mRNAs or gene models, and reducing the set of reference transcripts to only those found to overlap any of the input loci. With these options, Cuffcompare compares Cufflinks transfrags with a reference annotation and classifies them in different class codes, such as novel isoforms, intergenic or intronic transcripts. Isoform are only considered novel when a transfrag shared at least one splice junction with a reference transcript.

Expression quantification and differential expression analysis

After the assembly phase, the Cuffdiff tool (60) from Cufflinks was used to calculate expression values and to perform the differential expression analysis of the annotated genes and the newly predicted isoforms detected between H and L groups. Cuffdiff was run using the bias correction (-b option) to improve accuracy of transcript abundance estimates, and the 'rescue method' for multi-reads (-u option) to more accurately weight reads mapping to multiple locations in the genome. The remaining parameters were set as default.

The genes and new isoforms were filtered according to two criteria: i) a minimum mean group expression greater than 0.5 FPKM, ii) a fold change of the expression differences between H and L groups higher or equal to 1.5. The R package q-value (58) was employed to calculate the false-discovery rate. Finally, those genes with a p-value and a q-value ≤ 0.05 were considered as DE. Additionally, those novel isoforms

with a p-value ≤ 0.05 (equivalent to a q-value ≤ 0.25) were considered as DE.

RNAseq analysis using the CLC Genomics Workbench software

The whole RNAseq data analysis was also performed with the CLC Genomics Workbench using the same parameters as in the Tophat-Cufflinks analysis, to identify the overlapping between both methodologies. The strategies for filtering, mapping, transcript assembly and quantification of this software are based on Mortazavi et al., (41). Mapping parameters were set as default and the minimum and the maximum distance between pairs were defined as in the TopHat analysis. The broken pairs were excluded when counting the mapped reads to generate expression values. Expression data were normalized using the normalization by quantiles option and the differential expression analysis was performed assuming that the data has Gaussian distributions. Homogeneous variance was assumed and the normalized expression values were used for the differential expression analysis. Genes detected as differentially expressed (DE) by both methodologies were considered the most consistent.

Validation of differentially expressed genes and novel isoforms by q-PCR and sequencing

In order to evaluate the reliability and reproducibility of the results obtained in the RNA-Seq analyses, several new isoforms and gene expression measures were validate using complementary molecular analyses. Potentially novel isoforms identified by Cuffcompare were validated by cDNA sequencing. Among the DE novel isoforms, the ones identified for the *PLAC8* and *KIAA1462* genes were selected for validation. One primer pair was designed covering exons 1 to 3 of the *PLAC8* gene (reference sequence: ENSSSCT0000010120) and another one between exons 2 and 3 (reference sequence: ENSSSCG00000011020) of the *KIAA1462* gene (See supplementary table 2 for primer sequences). Different size of the PCR products will allow distinguishing the different isoforms. The PCRs reactions were performed in a final volume of 25 µl, containing 2.5 µl of cDNA, 1 unit of Taq polymerase

(Biotools), specific buffer, 2.5 mM of dNTPs and 0.5 µM of each primer. Thermocycling was carried out under the following conditions: 94ºC for 5 min, 40 cycles of 94ºC for 45 s, 60ºC for 45s and 72ºC for 45s, with a final extension of 72ºC for 10 min. The PCR reactions were carried out in a GeneAmp PCR System 9700 (Applied Biosystems, Warrington, UK). The PCR products were purified with the GFX™ PCR DNA purification kit (GE Healthcare, UK) according to the manufacturers' protocol. PCR products were sequenced with both forward and reverse primers using the 3100 BigDye® Terminator v3.1 Matrix Standard in a 3730 DNA Analyzer (Applied Biosystems Warrington, UK). The obtained sequences were edited and aligned using the EditSeq and MegAlign packages of the WinStar software for the comparison between the novel and the annotated isoforms in Ensembl.

In addition, to evaluate the reproducibility of the expression measures obtained, relative transcript quantification by RT-qPCR of seven differentially expressed genes (*BTG2*, *ACTA2*, *PRCP*, *FA2H*, *ADAMTS4*, *VAMP8* and *IRF1)* and two differentially expressed novel isoforms (*PLAC8* and *KIAA1462*) was performed in the eight animals. The reactions were performed in 384-well plates using the LightCycler®480 Real-Time PCR System (Roche Diagnostic, Mannheim, Germany). Four widely used presumable constitutively expressed reference genes, *GAPDH, B2M, TOP2B* and *ACTB*, were included in the plates to select the most suitable ones for expression data normalization. The GeNorm algorithm was used to identify the reference genes with the highest stability in this particular dataset. Real-time qPCR reactions were performed in a total volume of 20µl containing 2.5µl of cDNA (1/10 dilution), 10µl of Roche LightCycler mix and 0.5 µl of the specific primer pairs. All primer pairs used are detailed in Supplementary table 3. Standard PCR on cDNA was carried out to verify amplicon sizes. Cycling conditions were 95ºC for 10 min, followed by 45 cycles of 95ºC (15s) and 60ºC (1 min), when the fluorescence was acquired. A dissociation curve to test PCR specificity was generated by one cycle at 95ºC (15s), followed by 60ºC (20s) and ramped up to 95ºC with the

fluorescence acquired during the increase to 0.01ºC/s. All points and samples were run in triplicates as technical replicates and dissociation curves were analyzed for each individual replicate. Single peaks in the dissociation curves confirmed the specific amplification of the primer pairs and the absence of primer dimers. A non-template control, without cDNA, was included as negative control. PCR efficiency of each primer pair, shown in Supplementary table 3, was estimated by standard curve calculation using four points of cDNA serial dilutions (1, 1/2, 1/4 and 1/8) of a pool of the eight samples used. Data generated were analyzed with LightCycler 480 software (Roche) using the second derivative method (34). Mean Cp values were transformed to quantities using the comparative Cp method, setting the highest relative quantities for each gene to 1 (Quantity = $10^{-\Delta Cp/slope}$). Data normalization was carried out using the reference genes *GAPDH* and *ACTB* that showed the highest stability and efficiency values in these samples.

Pearson correlations were calculated between RT-qPCR and RNA-Seq expression data for the nine tested transcripts.

Gene functional classification, network and pathway analyses

A functional analysis of the genes expressed in hypothalamus was performed by analyzing pathways enrichment with the FatiGO browser from Babelomics 4.3 (http://babelomics.bioinfo.cipf.es) using the Reactome database. Similarly, a functional analysis of the differentially expressed genes between H and L groups was performed by analyzing GO enrichment with FatiGO using GO, KEGG and Reactome databases. This software provides an adjusted p-value according to the FDR procedure of Benjamini & Hochberg for the enrichment analyses. The functions and published literature of all relevant DE genes and isoforms have been analyzed using the NCBI (http://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/) databases. In addition, The Ingenuity Pathways Analysis software (IPA; Ingenuity Systems, www.ingenuity.com) was employed to generate networks that allow identifying the most significant

biological functions of the DE genes selected as most relevant (see results for details on selection).

In order to identify potential regulatory elements of the relevant genes, GeneCards database (www.genecards.org), SABiosciences' Text Mining Application (http://www.sabiosciences.com) and the UCSC Genome Browser (http://genome.ucsc.edu/) were used.

RESULTS AND DISCUSSION:

Characterization of the porcine hypothalamic transcriptome

A total of 1,028 million paired-end reads were obtained from the transcriptome sequencing of the eight hypothalamic samples. After trimming and filtering according to quality parameters, this number was reduced to 1,026 million reads (supplementary table 2). The TopHat pipeline mapped around 83% of those reads against the pig reference genome (build 10.2). The distribution of the gene expression values obtained with Cuffdiff for the 25,010 genes annotated in the pig genome is represented in Supplementary Figure 1. Only 4,762 of the annotated genes showed no expression in this dataset, while 16,024 genes were expressed in these samples with a mean expression value higher than 0.5 FPKM in both groups of animals. The number of expressed genes in this tissue was higher than found in other porcine tissues in previous studies. A total of 12,816 annotated genes were expressed in gonads (8), and around 11,000 genes in liver (25, 51). This result is in agreement with studies in mouse and human that have reported higher levels of RNA editing, associated to an increased in transcript diversity in brain compared to other pig tissues (6, 33). However, these differences could also be due to the use of different annotation of the pig genome, the higher sequencing depth obtained in the present study and the different filtering criteria set up in the different studies. Most of the genes showed a low or intermediate expression level between 0.5 and 100 FPKMs, and only a small proportion showed very high expression levels greater than 1,000 FPKMs, in agreement with the results obtained in other tissues (8, 51).

The functional enrichment analysis performed with Babelomics against the Reactome database revealed that genes expressed in the hypothalamus

are involved in a wide range of functions (Figure 2). As expected, among the strongest enriched pathways are several specific functions of the central nervous system such as axon guidance, nerve growth factor and synaptic transmission. In addition, it is worth mentioning that 35% of the enriched pathways identified are related to signaling pathways and therefore different types of growth factors signaling (platelet, epidermal, vascular, bone and fibroblast growth factors) are overrepresented, supporting the importance of the hypothalamus in signaling, growth and developmental processes. Also, diabetes appeared as the second most enriched pathway, showing the high importance of the hypothalamus in glucose and insulin metabolism and signaling. These results, together with the enrichment in lipid metabolic processes and energy metabolism integration, show the key role of this tissue for the traits of interest of the present study, supporting the decision to focus on the hypothalamus to investigate genes regulating growth and fatness. Moreover, previous studies based on the animal material used in the present study (Iberian x Landrace cross) have shown the relevance of the hypothalamic *LEPR* gene expression in growth and fatness (45, 49). An exonic polymorphism of this gene, *LEPR*c.1987C<T, is highly associated with growth, fatness and body composition (45). Differential hypothalamic expression of two different *LEPR* isoforms was found linked to this SNP (49). Unfortunately, although the obvious importance of this gene, it could not be analyzed in the present study because *LEPR* is incompletely annotated in the current genome annotation.

A total of 268,608 different transcripts were identified by Cufflinks in the eight hypothalamic samples. The classification of these transcripts performed with Cuffcompare is summarized in Table 1. A high percentage of the transcripts (27.8%) was annotated as intergenic transcripts illustrating the incompleteness of the available porcine genome annotation especially in annotation of non-coding RNA genes explaining this large number of novel intergenic transcripts. The high number of identified intergenic transcripts is in agreement with

previous studies in different porcine tissues: 36.1% of intergenic transcripts found in gonads (with the previous build 9 of the pig genome reference; 8) and 20.2% in liver (using the pig genome version 9.61 as reference; 51). Moreover, there is a large number of transcripts falling within a reference intron (15.3%) that indicate intron retention events, incorrect annotation of exons and even errors or missing prediction of isoforms. The later would be in agreement with the high number of potentially novel isoforms identified. A total of 58,927 potentially new isoforms was predicted by Cuffcompare, sharing at least one splice junction with a reference transcript and representing 21.9% of the total transcripts detected.

Differential expression analysis

Differentially expressed annotated genes

The number of paired-end reads obtained before and after trimming were similar in both groups, around 400 million reads for both H and L groups (Supplementary table 2). Similar percentages of mapped reads were also observed, 83.2% in the H group and 83.4% in the L group (Supplementary table 2). In addition, as shown in Supplementary Figure 1, global gene expression levels were equally distributed in both groups. Taking these into account, a larger part of the differences in gene expression observed between both groups can probably be attributted to the divergence of the two backcrossed breeds for growth and fatness traits. While the Landrace breed is characterized by a lean and fast growth the Iberian breed shows a higher food intake, higher proportion of fat and slower growth. The differential expression analysis of annotated genes performed with Cuffdiff revealed a total of 210 DE genes between H and L groups, according to the employed cut-off (mean expression ≥ 0.5 FPKM, fold change ≥ 1.5 and p-value and q-value ≤ 0.05). Among them, 17 genes showed highest expression in the H group and the remaining 193 DE genes showed highest expression in the L group. The fold change of these DE genes ranged from 1.65 to 18.19. The GO enrichment analysis performed by FatiGO (Babelomics) on the 210 DE genes indicated that several GO molecular functions

related to the traits of interest are overrepresented: peptidase activity (adjusted p-value = 1.1 $x10^{-2}$), growth factor binding (adjusted p-value = 5.0 $x10^{-3}$) and insulin-like growth factor binding (adjusted p-value = 4.9 $x10^{-2}$). Similarly, amongst the GO biological processes enriched in the DE genes, several processes related to growth and fatness traits could be identified: skeletal system development (adjusted p-value = 1.8 $x10^{-6}$), tissue development (adjusted p-value = 9.7 $x10^{-4}$), response to hormone stimulus (adjusted p-value = 4.7 $x10^{-8}$), transforming growth factor beta receptor signaling pathway (adjusted p-value = 1.5 $x10^{-5}$) and response to nutrient levels (adjusted p-value = 0.024). A selection of candidate genes was performed fitting two criteria: co-localization with previous QTLs identified in the same animal material and (/or) overlap with an independent differential expression analyses conducted with CLC Genomics software.

In total 15 of the 210 DE genes were located close (within 8 Mb from QTL confidence interval) or within these QTL regions (Table 2), highlighting relevant candidate genes. A deeper analysis of these 15 DE genes revealed that six of them are functionally linked to the traits of interest (*IRF1*, *ADAMTS4, FA2H, EGR-1, PMCH* and *MFAP5*; Table 2). The *IRF1* codes for a transcription factor (TF) involved in growth hormone regulation (21), the *ADAMTS4* gene is associated to an increase in beef marbling (32), the *FA2H* codes for a regulator of adipocyte differentiation and lipogenesis controlling body weight (18) and the TF *EGR-1* gene product is involved in the interaction between leptin and cholecystokinin hormones, which play important roles in regulating food intake, meal size and duration (30, 31). Moreover, a direct regulation of systemic cholesterol levels via regulation of cholesterol synthesis in the liver has also been reported for the *EGR-1* gene (14). *PMCH*, described as a hypothalamic-specific gene (48), codes for the melanin-concentrating hormone (*MCH*) precursor, a key regulator of food intake and metabolism (42). Finally, *MFAP5* gene expression has been correlated to changes in the amount of body fat, adiponectin and leptin levels, as well as to the expression of several genes associated with adipose tissue

development and differentiation such as *PPARG, CCND2* and *ADAM12* (61).

To focus on those genes with the highest evidence of being differentially expressed between both groups, an additional RNA-Seq data analysis was conducted with CLC Genomics Workbench software using the same parameters and reference genome as in the Tophat-Cufflinks analysis. There are several differences in the methodology used by both analyses (CLC user manual: http://www.clcsupport.com/clcgenomicsworkbench/current/; Cufflinks user manual: http://cufflinks.cbcb.umd.edu/howitworks.html). The CLC Genomics performs normalization by quantiles of the expression values and the Cufflinks uses the FPKM normalized values. Besides, the Cufflinks pipeline incorporates some options such as the multi-mapped read correction and the sequence bias correction that improve the estimation of the abundances. However, the greatest differences are in the approaches used to identify differentially expressed genes. The CLC Genomics assumes a Gaussian distribution of the data, whereas Cuffdiff "learns" the distribution from the input data providing a most accurate estimation. Additionally, the CLC Genomics performs a classic test based on the means and variances within groups and then calculates a t-test assuming "homogeneous" variances (in this case) and calculates a two-sided p-value, while the Cuffdiff is based in the log-fold-change, estimates the variance in the log-fold-change using simulated draws of the model of variance in expression for each group and calculates an empirical p-value based on these simulations. It is notable that the correlations between expression measures obtained by both RNA-Seq analysis methodologies, TopHat-Cufflinks and CLC Genomics were high and significant (the mean correlation was 0.88; p-value < 0.01). Moreover, the concordance correlation coefficient (CCC) (38) that provides a global indication of the reproducibility of gene expression studies was 0.98, indicating an almost perfect agreement between both RNA-Seq analysis methodologies. Even though the concordance in gene expression measures between both methodologies is high, as mentioned above, the strategies to analyze the differential expression are different.

Those genes that appeared as DE in both analyses were considered the most consistent. A total of 56 annotated genes were identified as DE by both approaches. Amongst them, 13 have previously been associated with growth and fatness traits in pig or other mammalian species and nine have known biological functions directly related to these traits: *PRCP, FA2H, ALDH2, IRF1, ANGPTL2, VAMP8, PLAC8, JUNB* and *TNFAIP3*. Two of these genes, *IRF1* and *FA2H*, are located within the QTL regions discussed above. The *PRCP* gene codes for a prolylcarboxipeptidase, a hypothalamic regulator of food intake and energy homeostasis in rodents (23, 64). The *JUNB* gene product has been described as a growth inhibiting protein and JUNB-deficient mice showed a retarded growth and a reduction of adipose tissue (50). The *PLAC8* gene codes for an important regulator of lipogenesis and adipocyte differentiation, being involved in the control of body weight (24). The *ALDH2* is a downstream target of *PPARG* signaling (43), and *ANGPTL2* and *VAMP8* genes are involved in insulin sensitivity and glucose metabolism (26, 28, 67). Finally, the *TNFAIP3* gene has been related to lipid and fatty acid metabolism in mice (5). For the remaining four genes (*BMP5, BTG2, ACTA2* and *ADAMTS4*), even though they do not have a known biological function directly related to growth and fatness, they have been previously linked to these traits through association analyses. The gene *ADAMTS4* mapped within a previously described QTL region and its association to the traits of interest was discussed above. The *BMP5* is located within a major QTL region identified in pig affecting carcass fat content and a SNP within this gene has been associated with fatness (55). Polymorphisms in the porcine *BTG2* and *ACTA2* genes have been previously associated to fat content (22, 39).

Identification of new isoforms differentially expressed

The existence of the new isoforms of the *PLAC8* and *KIAA1462* genes has been validated by cDNA sequencing. For the *PLAC8* gene, an isoform lacking exon 2 was confirmed. For the *KIAA1462* gene, a transcript containing the intronic region between exons 3 and 4 was also

confirmed. The validation of these two novel isoforms supports the reliability of the strategy used for new isoforms identification.

The differential expression analysis performed with the 58,927 new isoforms revealed a total of 50 DE isoforms between H and L groups. All of them showed a higher expression in the L group compared to the H group. These new isoforms are characterized by containing extra exons, intron retention events or exon skipping. The biological annotation showed that seven of them are splice forms of genes related to growth and fatness (*AEBP1*, *SETDB1*, *CD44, SLC44A1, PLAC8, BMP5* and *KIAA1462*). The *AEBP1* gene codes for the adipocyte enhancer-binding protein that modulates adiposity and energy homeostasis (52), the *CD44* seems to participate in the link between obesity and insulin resistance (27) and the *PLAC8* gene is a regulator of brown fat tissue adipocyte differentiation (24). The product of the gene *SLC44A1* is a choline transporter involved in lipid metabolism (35), and the *SETDB1* gene codes for a histone methyltransferase part of a repressor complex of *PPARG* that determines the differentiation of stem cells into adipocytes (17). Although the remaining two genes, *BMP5* and *KIAA1462* have not known function related to the traits of interest, they are located within QTL regions and contain SNPs previously associated with fat content and deposition, respectively (55, 65).

qPCR validation of differential expression analyses

Among the 56 DE genes identified both by Cuffdiff and CLC Genomics, seven were selected to be validated by qPCR. They included three DE genes located within previously detected QTL regions for growth and fatness (*FA2H*, *IRF1* and *ADAMTS4*) and four genes identified by both RNA-Seq analyses, which previously have been associated to the traits of interest (*PRCP*, *VAMP8*, *BTG2* and *ACTA2*). The expression values of the two new DE isoforms validated by sequencing (*PLAC8* and *KIAA1462*) were also selected for qPCR validation.

The expression values of the nine transcripts were measured by qPCR and normalized using *GAPDH* and *ACTB* as reference. The correlations

of the expression values obtained with qPCR, TopHat-Cufflinks and CLC Genomics are shown in Table 3. The correlations obtained between both RNA-Seq analyses and qPCR were highly significant, ranging from 0.67 (*PRCP* gene) to 0.98 (*PLAC8* new isoform). The average correlation between TopHat-Cufflinks and qPCR was 0.86 which is in agreement with the results from Roberts et al., (53) who reported a correlation of 0.81 between Cufflinks and qPCR expression data. It is not possible to calculate the CCC between qPCR and RNA-Seq methodologies due to several limitations: the limit number of validated genes, the non-random selection of these genes and the intrinsic differences in the expression measures of both types of methodologies (i.e. while all different transcripts of a gene are being measured in the RNA-Seq analyses only those covered by the primers designed are measured by qPCR).

Transcription factors (TF) related to growth and fatness traits

Several genes related to growth and fatness traits have been highlighted through the different approaches used (Table 4). A search for potential regulatory elements on this set of genes was conducted using GeneCards database and the SABiosciences' Text Mining Application. A total of 116 potential TFs regulators were found. Amongst them, TFs located within the QTL regions discussed above, were selected as powerful TF candidate genes. Six TFs are located within a QTL region: *IRF1* (QTL SSC2), *EGR-1* (QTL SSC2), *PBX1a* (QTL SSC4), *POU2F1* (QTL SSC4), *NR3C1* (QTL SSC2) and *NF-Yb* (QTL SSC5). Both *IRF1* and *EGR-1* are considered highly relevant as they also appeared differentially expressed between H and L groups. Both are TFs regulating a set of relevant DE genes (*BTG2*, *ACTA2*, *JUNB* and *CD44*) and are located within QTL regions associated to the traits of interest (SSC2: 148-152 Mb). Moreover, there is a high correlation between the expression of *IRF1* and *JUNB* (0.98), *IRF1* and *BTG2* (0.98) and *EGR-1* and *JUNB* (0.93). Remarkably, *EGR-1* is a potential regulator of *IRF1*, and their expression levels are highly correlated (0.91). Additionally, the *NR3C1*, a potential regulator of *SLC12A2, SELE, CPXM1, ACTA2* and *ALDH2*, also maps in the interval of

the same QTL region on SSC2 and it codes for a glucocorticoid receptor associated with obesity and eating disorders in humans (2). This gene also seems to be regulated by *EGR-1* and *IRF1*. Similarly, two of the identified TFs are located in the interval of SSC4 QTL (68-85 Mb), *PBX1a* and *POU2F1*. *PBX1a* regulates *MYL9* and *KIAA1462* DE transcripts and has been proposed as an adipocyte development regulator at multiple levels. It promotes the generation of adipocyte progenitors during embryogenesis, while favoring adipocyte progenitors proliferation and commitment to the adipocyte lineage in post-natal life (40). *POU2F1* regulates de expression of *ANO6* DE gene and has been previously associated to lipoprotein lipase transcriptional regulation (4). The final relevant TF is *NF-Yb,* located within the QTL region on SSC5 (69-83 Mb). It potentially controls the expression of *PRCP* DE gene, which has an important role in the regulation of food intake and energy homeostasis (23). The fact that some of these TFs are not DE conditional on the analyzed groups does not necessary mean that they are not related to these traits, as it is known that small expression variations of TFs could have great impact on its target genes expression.

Functional networks and pathway analyses

The IPA software was used to build a network with the most relevant DE genes and the TFs located in QTL regions and therefore potentially regulating DE genes. All these genes matched within four different networks (Table 5). Networks 1 and 2 were the most informative according to their scores, the number of discussed genes and the associated functions. In order to get further insights, networks 1 and 2 were merged and the resulting network is presented in Figure 3. In addition, the relevant canonical pathways related to the traits of interest, such as growth hormone, leptin and PPAR signaling were also included in the network. This overview of the network shows the important role of the abovementioned TFs in these processes, especially for *EGR-1, NR3C1* and *IRF1*. The *NR3C1* TF is linked with all the TFs of the network, *IRF1*, *PBX1*, *EGR-1* and *POU2F1*, highlighting a key regulatory role of

EGR-1 in growth and fatness. Finally, even though *JUNB* is not located in a QTL region, the high number of gene connections and its central location in the network, point out a relevant role of this TF in these processes.

Conclusions

The undertaken RNA-Seq approach has enabled us to explore the hypothalamic transcriptome of pigs divergent for growth and fatness traits. The transcriptome profile of the hypothalamus confirms its important role in the regulation of growth, fat accumulation and energy expenditure. The differential expression analyses performed with TopHat-Cufflinks revealed a total of 210 DE genes between H and L groups. Amongst them, 15 are located within QTL regions previously identified for these traits. An additional analysis conducted with CLC Genomics software supported the results obtained for 56 DE genes, of which 13 were biologically related to growth and fatness and three mapped within QTL regions. Finally, identification of new isoforms allowed us to identify a total of 50 DE novel isoforms, seven out of them related to the traits of interest. In summary, a set of 21 relevant differentially expressed genes could be highlighted. Additionally, six TFs were identified as potential regulators of the 21 DE genes. Among them, *EGR-1*, *IRF1* and *NR3C1* are the strongest candidate TF genes for growth and fatness according to their expression pattern, genomic location within QTL regions and network interactions. Overall, the results of the present study contribute to an improved knowledge of relevant metabolic routes for growth and fatness traits, not only relevant for pig production, but also in relation to human obesity and related diseases. The next step towards the search of the causal genetic variation would be the analysis of polymorphisms in the promoter regions of the highlighted genes and even in the transcription factors potentially responsible of the expression differences found in the current study.

ACKNOWLEDGEMENTS

We want to thanks Rita Benítez, Yolanda Nuñez and Fabián García for their technical support. Moreover we want to acknowledge Luis Silió and Noelia Ibáñez-Escriche for their help in the revision of the manuscript.

GRANTS

This work was funded by Ministerio de Ciencia e Innovación (MICINN) project AGL2011-29821-C02. DPM was funded by a Formación de Personal Investigador (FPI) PhD grant from the Spanish Ministerio de Ciencia e Innovación (BES-2009-025417). OM and MAMG was supported by European Research Council grant no. ERC-2009-AdG:249894.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTION

Author contributions: AIF and JMF conducted the conception and design of research; DPM and EA performed experiments; DPM, OM, MAMG and AIF analyzed data; DPM, OM and AIF interpreted results of the experiments; MCR and JLN arranged the phenotypic collection and sampling; DPM and AIF drafted the first version of the manuscript.

REFERENCES:

1. **Barb CR, Hausman GJ, Rekaya R, Lents CA, Lkhagvadorj S, Qu L, Cai W, Couture OP, Anderson LL, Dekkers JC, Tuggle CK.** Gene expression in hypothalamus, liver and adipose tissues and food intake response to melanocortin-4 receptor (MC4R) agonist in pigs expressing MC4R mutations. Physiol Genomics. 41:3: 254-268, 2010.
2. **Cellini E, Castellini G, Ricca V, Bagnoli S, Tedde A, Rotella CM, Faravelli C, Sorbi S, Nacmias B.** Glucocorticoid receptor gene polymorphisms in Italian patients with eating disorders and obesity. Psychiatr Genet 20(6):282-8, 2010.
3. **Chen C, Ai H, Ren J, Li W, Li P, Qiao R, Ouyang J, Yang M, Ma J, Huang L.** A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. BMC Genomics 12:448, 2011.
4. **Currie RA, Eckel RH.** Characterization of a high affinity octamer transcription factor binding site in the

human lipoprotein lipase promoter. Arch Biochem Biophys 298(2):630-9, 1992.

5. **Damrauer SM, Studer P, da Silva CG, Longo CR, Ramsey HE, Csizmadia E, Shrikhande GV, Scali ST, Libermann TA, Bhasin MK, Ferran C.** A20 modulates lipid metabolism and energy production to promote liver regeneration. PLoS One 6(3), 2011.
6. **Danecek P, Nellåker C, McIntyre RE, Buendia-Buendia JE, Bumpstead S, Ponting CP, Flint J, Durbin R, Keane TM, Adams DJ.** High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol 13(4):26, 2012.
7. **Estellé J, Pérez-Enciso M, Mercadé A, Varona L, Alves E, Sánchez A, Folch JM.** Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross. Anim Genet 37(6):589-91, 2006.
8. **Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, Notredame C, Pérez-Enciso M.** Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. BMC Genomics 12:552, 2011.
9. **Fernández AI, Pérez-Montarelo D, Barragán C, Ramayo-Caldas Y, Ibáñez-Escriche N, Castelló A, Noguera JL, Silió L, Folch JM, Rodríguez MC.** Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. BMC Genet 13:41, 2012.
10. **Ferraz AL, Ojeda A, López-Béjar M, Fernandes LT, Castelló A, Folch JM, Pérez-Enciso M.** Transcriptome architecture across tissues in the pig. BMC Genomics 16; 9:173, 2008.
11. **Fortes MR, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, Hawken RJ, DeAtley KL, Peters SO, Silver GA, Rincon G, Medrano JF, Islas-Trejo A, Thomas MG.** Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome

information, and transcription factors. J Anim Sci 90(9):2894-906, 2012.

12. **Gorodkin J, Cirera S, Hedegaard J, Gilchrist MJ, Panitz F, Jørgensen C, Scheibye-Knudsen K, Arvin T, Lumholdt S, Sawera M, Green T, Nielsen BJ, Havgaard JH, Rosenkilde C, Wang J, Li H, Li R, Liu B, Hu S, Dong W, Li W, Yu J, Wang J, Staefeldt HH, Wernersson R, Madsen LB, Thomsen B, Hornshøj H, Bujie Z, Wang X, Wang X, Bolund L, Brunak S, Yang H, Bendixen C, Fredholm M.** Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. Genome Biol 8(4):R45, 2007.
13. **Goff L, Trapnell C, Kelley D**. CummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2.0.0. 2012.
14. **Gokey NG, Lopez-Anido C, Gillian-Daniel AL, Svaren J.** Early growth response 1 (Egr1) regulates cholesterol biosynthetic gene expression. J Biol Chem 286(34):29501-10, 2011.
15. **Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LA, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi SH, Chow W, Clark RC, Clee C, Crooijmans RP, Dawson HD, Dehais P, De Sapio F, Dibbits B, Drou N, Du ZQ, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GJ, Fowler KE, Fredholm M, Fritz E, Gilbert JG, Giuffra E, Gorodkin J, Griffin DK, Harrow JL, Hayward A, Howe K, Hu ZL, Humphray SJ, Hunt T, Hornshøj H, Jeon JT, Jern P, Jones M, Jurka J, Kanamori H, Kapetanovic R, Kim J, Kim**

JH, Kim KW, Kim TH, Larson G, Lee K, Lee KT, Leggett R, Lewin HA, Li Y, Liu W, Loveland JE, Lu Y, Lunney JK, Ma J, Madsen O, Mann K, Matthews L, McLaren S, Morozumi T, Murtaugh MP, Narayan J, Nguyen DT, Ni P, Oh SJ, Onteru S, Panitz F, Park EW, Park HS, Pascal G, Paudel Y, Perez-Enciso M, Ramirez-Gonzalez R, Reecy JM, Rodriguez-Zas S, Rohrer GA, Rund L, Sang Y, Schachtschneider K, Schraiber JG, Schwartz J, Scobie L, Scott C, Searle S, Servin B, Southey BR, Sperber G, Stadler P, Sweedler JV, Tafer H, Thomsen B, Wali R, Wang J, Wang J, White S, Xu X, Yerle M, Zhang G, Zhang J, Zhang J, Zhao S, Rogers J, Churcher C, Schook LB. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491(7424):393-8, 2012.

16. **Gunawan A, Sahadevan S, Neuhoff C, Große-Brinkhaus C, Gad A, Frieden L, Tesfaye D, Tholen E, Looft C, Uddin MJ, Schellander K, Cinar MU.** RNA deep sequencing reveals novel candidate genes and polymorphisms in boar testis and liver tissues with divergent androstenone levels. PLoS One 16;8(5), 2013.
17. **Günther T, Schüle R.** Fat or bone? A non-canonical decision. Nat Cell Biol 9(11):1229-31, 2007.
18. **Guo L, Zhou D, Pryse KM, Okunade AL, Su X.** Fatty acid 2-hydroxylase mediates diffusional mobility of Raft-associated lipids, GLUT4 level, and lipogenesis in 3T3-L1 adipocytes. J Biol Chem 285(33):25438-47, 2010.
19. **Hillebrand JJ, de Wied D, Adan RA.** Neuropeptides, food intake and body weight regulation: a hypothalamic focus. Peptides 23(12):2283-306, 2002.
20. **Hu ZH, Park CA, Wu XL, Reecy JM.** Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids

Research, 41 (D1): D871-D879, 2013.

21. **Hu J, Lou D, Carow B, Winerdal ME, Rottenberg M, Wikström AC, Norstedt G, Winqvist O.** LPS regulates SOCS2 transcription in a type I interferon dependent autocrine-paracrine loop. PLoS One 7(1):e30166, 2012.
22. **Huang JS, Xiong YZ, Deng CY.** Cloning and expression analysis of porcine ACTA2 gene and its association with production traits. Yi Chuan 31(5):489-94, 2009.
23. **Jeong JK, Szabo G, Kelly K, Diano S.** Prolyl carboxypeptidase regulates energy expenditure and the thyroid axis. Endocrinology 153(2):683-9, 2012.
24. **Jimenez-Preitner M, Berney X, Uldry M, Vitali A, Cinti S, Ledford JG, Thorens B.** Plac8 is an inducer of C/EBPβ required for brown fat differentiation, thermoregulation, and control of body weight. Cell Metab 14(5):658-70, 2011.
25. **Jung WY, Kwon SG, Son M, Cho ES, Lee Y, Kim JH, Kim BW, Park da H, Hwang JH, Kim TW, Park HC, Park BY, Choi JS, Cho KK, Chung KH, Song YM, Kim IS, Jin SK, Kim DH, Lee SW, Lee KW, Bang WY, Kim CW.** RNA-Seq approach for genetic improvement of meat quality in pig and evolutionary insight into the substrate specificity of animal carbonyl reductases. PLoS One 7(9), 2012.
26. **Kadomatsu T, Tabata M, Oike Y.** Angiopoietin-like proteins: emerging targets for treatment of obesity and related metabolic diseases. FEBS J 278(4):559-64, 2011.
27. **Kang HS, Liao G, DeGraff LM, Gerrish K, Bortner CD, Garantziotis S, Jetten AM.** CD44 plays a critical role in regulating diet-induced adipose inflammation, hepatic steatosis, and insulin resistance**.** PLoS One 8(3):e58417, 2013.
28. **Kitazawa M, Nagano M, Masumoto KH, Shigeyoshi Y, Natsume T, Hashimoto S.** Angiopoietin-like 2, a circadian gene, improves type 2 diabetes through

potentiation of insulin sensitivity in mice adipocytes. Endocrinology 152(7):2558-67, 2011.

29. **Kogelman LJ, Kadarmideen HN, Mark T, Karlskov-Mortensen P, Bruun CS, Cirera S, Jacobsen MJ, Jørgensen CB, Fredholm M.** An f2 pig resource population as a model for genetic studies of obesity and obesity-related diseases in humans: design and genetic parameters. Front Genet 4:29.
30. **de Lartigue G, Lur G, Dimaline R, Varro A, Raybould H, Dockray GJ.** EGR1 Is a target for cooperative interactions between cholecystokinin and leptin, and inhibition by ghrelin, in vagal afferent neurons. Endocrinology 151(8):3589-99, 2010.
31. **de Lartigue G, Barbier de la Serre C, Espero E, Lee J, Raybould HE.** Leptin resistance in vagal afferent neurons inhibits cholecystokinin signaling and satiation in diet induced obese rats. PLoS One 7(3):e32967, 2012.
32. **Lee SH, Gondro C, van der Werf J, Kim NK, Lim DJ, Park EW, Oh SJ, Gibson JP, Thompson JM.** Use of a bovine genome array to identify new biological pathways for beef marbling in Hanwoo (Korean Cattle). BMC Genomics 11:623, 2010.
33. **Li JB, Levanon EY, Yoon JK, Aach J, Xie B, Leproust E, Zhang K, Gao Y, Church GM.** Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science 324(5931):1210-3, 2009.
34. **Luu-The V, Paquet N, Calvo E, Cumps J.** Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. BioTechniques 38: 287-293, 2005.
35. **Michel V, Singh RK, Bakovic M.** The impact of choline availability on muscle lipid metabolism. Food Funct 2(1):53-62, 2011.

36. **Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundström K, Reinsch N, Gellin J, Kalm E, Roy PL, Chardon P, Andersson L.** A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science 19:1248-1251, 2000.
37. **Miller LC, Fleming D, Arbogast A, Bayles DO, Guo B, Lager KM, Henningson JN, Schlink SN, Yang HC, Faaberg KS, Kehrli ME Jr.** Analysis of the swine tracheobronchial lymph node transcriptomic response to infection with a Chinese highly pathogenic strain of porcine reproductive and respiratory syndrome virus. BMC Vet Res 8:208, 2012.
38. **Miron M, Woody OZ, Marcil A, Murie C, Sladek R, Nadon R.** A methodology for global validation of microarray experiments. BMC Bioinformatics 7:333, 2006.
39. **Mo XY, Lan J, Jiao QZ, Xiong YZ, Zuo B, Li FE, Xu DQ, Lei MG.** Molecular characterization, expression pattern and association analysis of the porcine BTG2 gene. Mol Biol Rep Oct;38(7):4389-96, 2011.
40. **Monteiro MC, Sanyal M, Cleary ML, Sengenès C, Bouloumié A, Dani C, Billon N.** PBX1: a novel stage-specific regulator of adipocyte development. Stem Cells 29(11):1837-48, 2011.
41. **Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B.** Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621-8, 2008.
42. **Mul JD, la Fleur SE, Toonen PW, Afrasiab-Middelman A, Binnekade R, Schetters D, Verheij MM, Sears RM, Homberg JR, Schoffelmeer AN, Adan RA, DiLeone RJ, De Vries TJ, Cuppen E.** Chronic loss of melanin-concentrating hormone affects motivational aspects of feeding in the rat. PLoS One 6(5):e19600, 2011.

43. **Okuno M, Arimoto E, Nishizuka M, Nishihara T, Imagawa M.** Isolation of up- or down-regulated genes in PPARgamma-expressing NIH-3T3 cells during differentiation into adipocytes. FEBS Lett 519(1-3):108-12, 2002.
44. **Ovilo C, Pérez-Enciso M, Barragán C, Clop A, Rodríquez C, Oliver MA, Toro MA, Noruera JL.** A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6. Mamm Genome 11(4):344-6, 2000.
45. **Ovilo C, Fernández A, Fernández AI, Folch JM, Varona L, Benítez R, Nuñez Y, Rodríguez C, Silió L.** Hypothalamic expression of porcine leptin receptor (LEPR), neuropeptide Y (NPY), and cocaine- and amphetamine-regulated transcript (CART) genes is influenced by LEPR genotype. Mamm Genome 21(11-12):583-91, 2010.
46. **Pérez-Enciso M, Varona L.** Quantitative trait loci mapping in F(2) crosses between outbred lines. Genetics 155(1):391-405, 2000.
47. **Pérez-Enciso M, Ferraz AL, Ojeda A, López-Béjar M.** Impact of breed and sex on porcine endocrine transcriptome: a bayesian biometrical analysis. BMC Genomics 10:89, 2009.
48. **Pérez-Montarelo D, Hudson NJ, Fernández AI, Ramayo-Caldas Y, Dalrymple BP, Reverter A**. Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. PLoS One 7(9):e46159, 2012.
49. **Pérez-Montarelo D, Fernández A, Barragán C, Noguera JL, Folch JM, Rodríguez MC, Ovilo C, Silió L, Fernández AI.** Transcriptional Characterization of Porcine Leptin and Leptin Receptor Genes. PLoS One 8(6):e66398, 2013.
50. **Pinent M, Prokesch A, Hackl H, Voshol PJ, Klatzer A, Walenta E, Panzenboeck U, Kenner L, Trajanoski Z, Hoefler G, Bogner-Strauss JG.** Adipose triglyceride lipase

and hormone-sensitive lipase are involved in fat loss in JunB-deficient mice. Endocrinology 152(7):2678-89, 2011.

51. **Ramayo-Caldas Y, Mach N, Esteve-Codina A, Corominas J, Castelló A, Ballester M, Estellé J, Ibáñez-Escriche N, Fernández AI, Pérez-Enciso M, Folch JM.** Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics 13:547, 2012.
52. **Ro HS, Zhang L, Majdalawieh A, Kim SW, Wu X, Lyons PJ, Webber C, Ma H, Reidy SP, Boudreau A, Miller JR, Mitchell P, McLeod RS.** Adipocyte enhancer-binding protein 1 modulates adiposity and energy homeostasis 15(2):288-302, 2007.
53. **Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L.** Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol 12(3):R22, 2011.
54. **Samborski A, Graf A, Krebs S, Kessler B, Bauersachs S.** Deep sequencing of the porcine endometrial transcriptome on day 14 of pregnancy. Biol Reprod 88(4):84, 2013.
55. **Shao GC, Luo LF, Jiang SW, Deng CY, Xiong YZ, Li FE.** A C/T mutation in microRNA target sites in BMP5 gene is potentially associated with fatness in pigs. Meat Sci 87(3):299-303, 2011.
56. **Simpson KA, Martin NM, Bloom SR.** Hypothalamic regulation of food intake and clinical therapeutic applications. Arq Bras Endocrinol Metabol 53(2):120-8, 2009.
57. **Spurlock ME, Gabler NK**. The development of porcine models of obesity and the metabolic syndrome. J Nutr 138(2):397-402, 2008.
58. **Storey JD, Tibshirani R.** Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100(16):9440-5, 2003.
59. **Trapnell C, Pachter L, Salzberg SL.** TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105-11, 2009.

60. **Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L.** Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562-78, 2012.
61. **Vaittinen M, Kolehmainen M, Schwab U, Uusitupa M, Pulkkinen L.** Microfibrillar-associated protein 5 is linked with markers of obesity-related extracellular matrix remodeling and inflammation. Nutr Diabetes 1:e15, 2011.
62. **Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L.** A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425:832-836, 2003.
63. **Varona L, Ovilo C, Clop A, Noguera JL, Pérez-Enciso M, Coll A, Folch JM, Barragán C, Toro MA, Babot D, Sánchez A.** QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects. Genet Res 80(2):145-54, 2002.
64. **Wallingford N, Perroud B, Gao Q, Coppola A, Gyengesi E, Liu ZW, Gao XB, Diament A, Haus KA, Shariat-Madar Z, Mahdi F, Wardlaw SL, Schmaier AH, Warden CH, Diano S.** Prolylcarboxypeptidase regulates food intake by inactivating alpha-MSH in rodents. J Clin Invest 119(8):2291-303, 2009.
65. **Xiao Q, Wu XL, Michal JJ, Reeves JJ, Busboom JR, Thorgaard GH, Jiang Z.** A novel nuclear-encoded mitochondrial poly(A) polymerase PAPD1 is a potential candidate gene for the extreme obesity related phenotypes in mammals. Int J Biol Sci 2(4):171-8, 2006.
66. **Yang B, Bassols A, Saco Y, Pérez-Enciso M.** Association between plasma metabolites and gene expression profiles in five porcine endocrine tissues. Genet Sel Evol 43:28, 2011.

67. **Zong H, Wang CC, Vaitheesvaran B, Kurland IJ, Hong W, Pessin JE.** Enhanced energy expenditure, glucose utilization, and insulin sensitivity in VAMP8 null mice. Diabetes 60(1):30-8, 2011.

TABLES:

Table 1. Classification of the transcripts identified in the hypothalamic samples in relation to the Ensembl annotated pig genes.

Cuffcompare class	**Number transcripts**	**% transcripts**
Complete match of intron chain	8,702	3.2
Multiple classifications	33,965	12.6
Contained in the reference	22,923	8.5
Possible pre-mRNA fragment	9,122	3.4
Transcript falling within a reference intron	41,008	15.3
Potentially novel isoforms	58,927	21.9
Generic overlap with a reference transcript	8,196	3.0
Possible polymerase run-on fragment	7,763	2.9
Intergenic transcript	74,845	27.8
Exonic overlap on the opposite strand	3,157	1.1
TOTAL	**268,608**	**100**

Table 2. Differentially expressed (DE) genes matching QTL regions previously detected for the traits of interest in the same animal material (8) and their respective position.

Chr.	DE Genes	Position (Mb)	QTL CI (Mb)	Associated Trait
2	*IRF1**	140	150-158	Mean weight of shoulders, weight of bone-in loins
	*EGR-1**	146		
4	*ADAMTS4**	97	68-85	Mean weight of shoulders, weight of bone-in loins
	C8ORF46	74		
	SELE	88		
4	*HFM1*	137	129-134	Backfat thickness at 75 kg
5	*SLC6A13*	69	69-83	Backfat thickness at slaughter
	*MFAP5**	65		
	*PMCH**	85		
6	*FA2H**	128	134-145	Backfat thickness at slaughter
14	*CPXM2*	145	149-153	Backfat thickness at 75 kg, Backfat thickness at slaughter
17	*MYL9*	45	39-42	Backfat thickness at 75 kg
	TGM2	46		
	PROCR	43		
	CPXM1	37		

* DE genes with biological functions related to the traits of interest. Chr, chromosome; CI, confidence interval.

Table 3. Pearson correlation between the expression values obtained from qPCR, TopHat-Cufflinks (TC) and CLC Genomics (CLC) for nine transcripts/genes.

Gene	qPCR-TC	*p*-value	qPCR-CLC	*p*-value
Annotated genes				
PRCP	0.69	$2x10^{-2}$	0.67	$3x10^{-2}$
FA2H	0.74	$1x10^{-2}$	0.72	$2x10^{-2}$
IRF1	0.92	$4x10^{-4}$	0.91	$6x10^{-4}$
VAMP8	0.88	$1x10^{-3}$	0.88	$1x10^{-3}$
BTG2	0.94	$1x10^{-4}$	0.92	$4x10^{-4}$
ACTA2	0.97	$1x10^{-5}$	0.96	$5x10^{-5}$
ADAMTS4	0.80	$7x10^{-3}$	0.79	$8x10^{-3}$
New isoforms				
PLAC8	0.98	$3x10^{-6}$	-	-
KIAA1462	0.84	$4x10^{-3}$	-	-

Table 4. Relevant differentially expressed genes identified from the comparison between high and low groups.

Ensembl ID	Gene ID	Fold Change	*q*-value
ENSSSCG00000010447	*ACTA2*	3.15	$3.6x10^{-4}$
ENSSSCG00000006359	*ADAMTS4*	1.72	$4.9x10^{-4}$
ENSSSCG00000016754	*AEBP1*	4.34	$3.0x10^{-1}$
ENSSSCG00000009889	*ALDH2*	1.79	$3.8x10^{-2}$
ENSSSCG00000005608	*ANGPTL2*	2.48	$1.2x10^{-2}$
ENSSSCG00000001478	*BMP5*	2.31	$4.7x10^{-2}$
ENSSSCG00000028322	*BTG2*	3.21	$9.0x10^{-5}$
ENSSSCG00000013297	*CD44*	4.87	$3.1x10^{-1}$
ENSSSCG00000014336	*EGR-1*	1.9	$3.7x10^{-2}$
ENSSSCG00000002718	*FA2H*	1.79	$4.0x10^{-2}$
ENSSSCG00000014277	*IRF1*	1.98	$3.1x10^{-2}$
ENSSSCG00000013735	*JUNB*	3.28	$2.3x10^{-3}$
ENSSSCG00000011020	*KIAA1462*	1.9	$3.1x10^{-1}$
ENSSSCG00000000666	*MFAP5*	3.77	$4.8x10^{-2}$
ENSSSCG00000009240	*PLAC8*	2.93	$3.1x10^{-1}$
ENSSSCG00000000858	*PMCH*	1.83	$4.6x10^{-2}$
ENSSSCG00000014899	*PRCP*	1.8	$4.1x10^{-2}$
ENSSSCG00000006645	*SETDB1*	2.79	$3.1x10^{-1}$
ENSSSCG00000014255	*SLC12A2*	3.17	$3.1x10^{-1}$
ENSSSCG00000005425	*SLC44A1*	2.01	$3.1x10^{-1}$
ENSSSCG00000022820	*VAMP8*	1.81	$4.7x10^{-2}$

Table 5. Networks built with the relevant DE genes and TFs (in bold) by the IPA software.

ID	Molecules in Network	Score	Focus Molecules	Top Functions
1	***ACTA2***, ***AEBP1***, ***ALDH2***, ***ANGPTL2***, *Ap1, BRC,* ***BTG2***, *CaMKII, Cbp/p300, CD3,* ***CD44***, *CD82, Cdc2, Cg, Creb,* ***EGR1***, *ERK, ERK1/2, FSH, GNRH, HIST3H3, IL1,* ***JUNB***, *Lh, Mapk, Mek, P38 MAPK, Pdgf, PDGF BB,* ***PMCH***, ***SLC12A2***, *TCF, Tgf β,* ***VAMP8***, *Vegf*	28	11	Cancer, neurological disease, endocrine system disorders
2	*Akt, ALOX12/15, AMIGO2, Atf,* ***BMP5***, *C8orf4, CARD14, Ck2, ESR1, FASTKD1, FGD4, PTK2, HIS4, IL12, IP6K3,* ***IRF1***, *Jnk, NFkB,* ***NR3C1***, *Oc1/2, Orm,* ***PBX1***, *Pka, Pkc, Ras* ***PLAC8***, ***POU2F1***, *RDM1, RELT, POLR2A, Taok2, TCR, Tnfrsf22/Tnfrsf23, vitK1, ZNF675*	12	6	Cellular development, growth and proliferation, organ morphology
3	***ADAMTS4***, *AKAP17A, ANO2, CDS2, CEP57L1, CHPT1, CHUK, CTC1, FAM184B, FAM49B, FAM92A1, FLYWCH1, KBTBD8, L2HGDH, METTL9, METTL2B,* ***MFAP5***, ***NFYB***, ***PRCP***, *PXMP4, RIMBP2, SETDB1,* ***SLC44A1***, *SNX10, TM7SF3, TMEM39A, TTC39B, TYW1, UBC, UBE3B, ZNF468, ZNF484, ZNF610, ZNF611, ZNF765*	12	6	Cardiovascular disease, connective tissue disorders, developmental disorder
4	C18-ceramide, cholesterol, DC16-ceramide, *Epgn,* ***FA2H***, fatty acid, lipid, *miR-124-3p, SBDS, UBQLN4*	2	1	Ophthalmic disease, hair, skin and organ development

FIGURES

Figure 1. Graphical representation of the first and second principal components of the PCA summarizing the phenotypical variation of the four traits related to growth and fatness. Animals assigned to the High (H) group are highlighted with white circles, and those to the Low (L) group with grey circles.

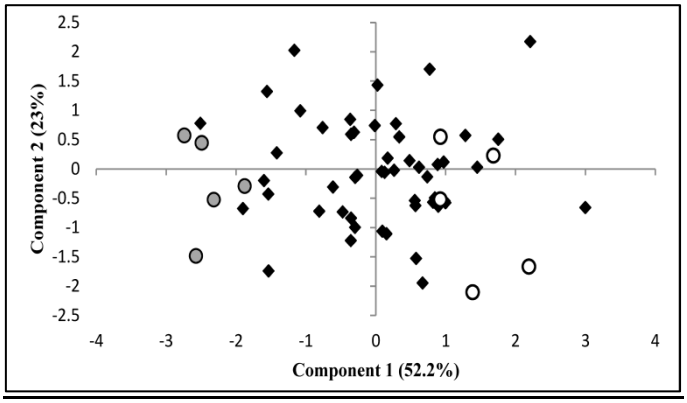


Figure 2. Reactome pathways enriched in the set of genes expressed in the hypothalamus samples analysed.

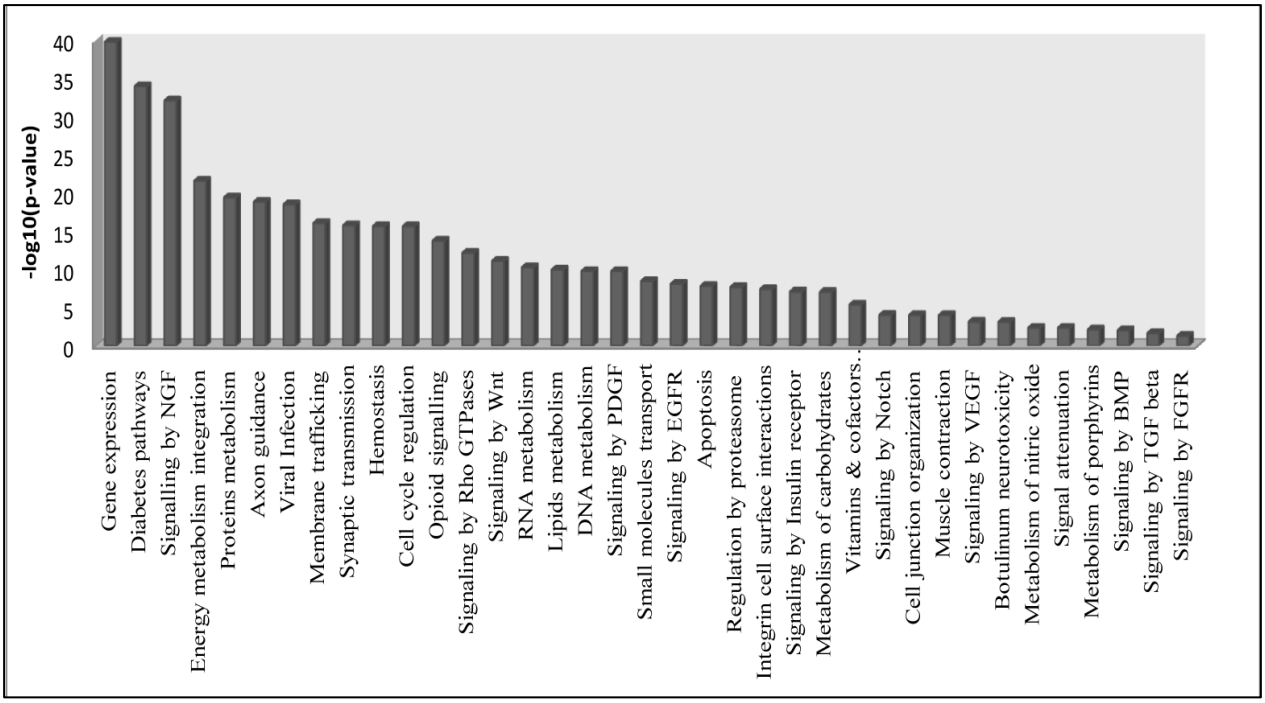
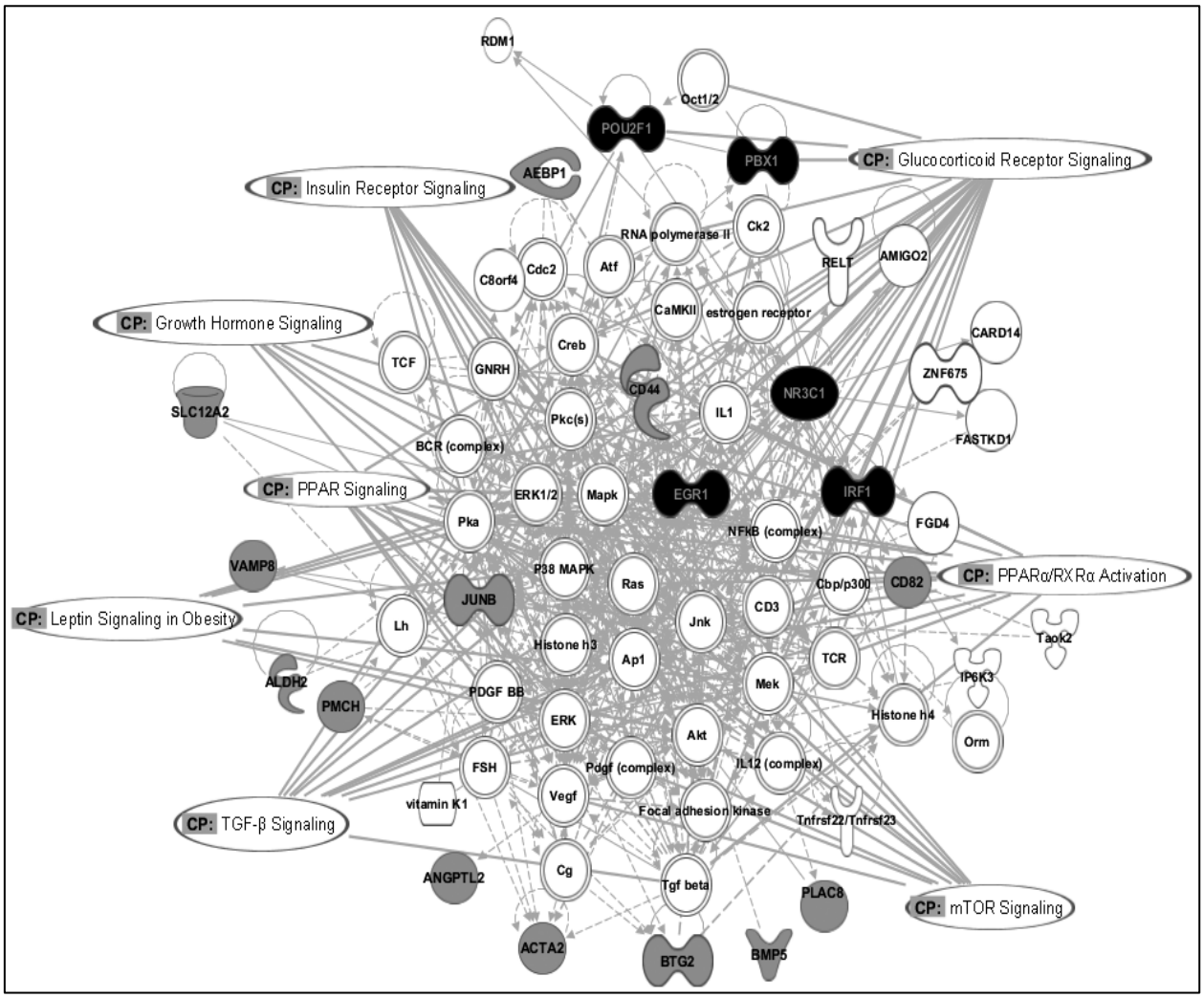


Figure 3. Graphical representation of merged networks 1 and 2 generated by the IPA software. The relevant DE genes are colored in grey, TFs in black and the remaining genes added to build the network in white. Overrepresented canonical pathways related to the traits of interest were also overlaid on the network.



ARTÍCULO III

Joint effects of porcine *leptin* and *leptin receptor* polymorphisms on productivity and quality traits

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION

doi: 10.1111/j.1365-2052.2012.02338.x

Joint effects of porcine *leptin* and *leptin receptor* polymorphisms on productivity and quality traits

D. Pérez-Montarelo*, A. Fernández*, J. M. Folch†, R. N. Pena‡, C. Óvilo*, C. Rodríguez*, L. Silió* and A. I. Fernández*

*Department Mejora Genética Animal, INIA, 28040, Madrid, Spain. †Department de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain. ‡IRTA, Genètica i Millora Animal, 25198, Lleida, Spain

Summary

Leptin signalling plays a fundamental role in growth, fatness and body composition. The aim of this study was to investigate the porcine *LEP* gene sequence in an Iberian × Landrace experimental cross to identify polymorphisms associated with productivity and quality traits. Because of the documented effects on these traits of the *LEPR* c.1987C>T polymorphism, the *LEP* and *LEPR* c.1987C>T polymorphisms and their interactions have been jointly investigated. The *LEP* gene sequencing has allowed the identification of 39 polymorphisms, eight of which are novel. Three intronic SNPs, *LEP g.1382C>T*, *LEP g.1387C>T* and *LEP g.1723A>G*, have been genotyped, and association analyses have been carried out. Analyses of *LEP* g.1387C>T, fully linked to *LEP* g.1382C>T, have revealed additive effects on live and carcass weights and dominant effects on several backfat thickness measurements. Novel effects of both *LEP* and *LEPR* polymorphisms on fatty acid composition in subcutaneous fat have been detected, probably mediated through the effects on fatness. The results reported here suggest that the *T* alleles of both *LEP* g.1387C>T and *LEPR* c.1987C>T, which are fixed in the Iberian pigs, would lead to an increase in growth, fatness and saturated fatty acid content in fat, which could be explained by an increased feed intake.

Keywords leptin, leptin receptor, productive traits, porcine, quality traits.

Leptin signalling plays fundamental roles in food intake and energy expenditure, thereby influencing body mass and composition (Barb *et al.* 2001). Leptin is mainly synthesized by white adipocytes and secreted into the bloodstream. In the brain, it regulates, through its receptor, the expression of neurotransmitters that have a role in reducing feed intake and increasing energy expenditure, thereby reducing adipose mass. Leptin also reaches some peripheral tissues like pancreas, liver or muscle, regulating its metabolism (Houseknecht *et al.* 1998). Despite the relevance of the *leptin* (*LEP*) gene in the pig (Wylie 2011), which has been mapped on SSC18, previous studies have shown conflicting results regarding the association of polymorphisms in this gene with production and reproductive traits (Jiang & Gibson 1999; Kennes *et al.* 2001; Chen *et al.* 2004; De Oliveira Peixoto *et al.* 2006; Amills *et al.* 2008; Switonski *et al.* 2010). The porcine *leptin receptor* gene (*LEPR*), located in SSC6, has been also widely considered a candidate gene underlying a QTL for growth, fatness and body composition detected in an Iberian × Landrace experimental cross IBMAP (Óvilo *et al.* 2002, 2005). These studies revealed a highly significant association of the *LEPR* c.987C>T polymorphism (previously described as c.2002C>T) with growth and fatness, and these effects have been confirmed in other porcine populations (Muñoz *et al.* 2009, 2011; Rodríguez *et al.* 2010). Recently, functional studies have revealed differences in the *LEPR* mRNA levels in the hypothalamus conditional on the *LEPR* c.1987C>T genotype (Óvilo *et al.* 2010).

In the current study, we have analysed the porcine *LEP* sequence variants in the IBMAP cross to investigate their association with growth, fatness, fatty acid composition and conformation traits. Moreover, we have investigated the joint effects of *LEPR* and *LEP* polymorphisms on the analysed traits, and their interaction.

Address for correspondence

A.I. Fernández, Dpto. Mejora Genética Animal, INIA, Ctra. Coruña Km 7.5, 28040 Madrid, Spain.
E-mail: avila@inia.es

Accepted for publication 8 December 2011

The growth, fatness, conformation and fatty acid composition records used in this study (Table S1) belong to the F_2 and F_3 generations of the IBMAP experimental population (Óvilo *et al.* 2000). Genomic DNA from some parental animals, five Landrace females and three Iberian Guadyerbas males, was used for *LEP* sequencing and polymorphism detection. Nine primer pairs (Table S2) were designed (from GenBank sequence AJ865080.1) to amplify the almost complete *LEP* gene. PCR conditions are described in Table S2. PCR products were sequenced in a 3730 DNA Analyzer (Applied Biosystems), and sequences were edited and aligned using WINSTAR software. A total fragment of 5031 bp, covering positions 849–5880 of the reference *LEP* sequence, was successfully sequenced. The fragment covered the complete coding region of the gene (second and third exons), intron 2, and a portion of intron 1 and the 3′ UTR. The sequence comparison revealed 39 polymorphisms, eight of which were novel (Table S3), all located in introns except one, the *LEP* g.3901C>T polymorphism, which is a synonymous SNP located in the third exon (Figure S1). We selected three SNPs, *LEP g.1382C>T*, *LEP g.1387C>T* and *LEP g.1723A>G*, for genotyping and association analyses. Selection of the SNPs was based on their potential informativity, according to the genotypic frequencies in the parental pigs, their genotyping feasibility and their ability to represent most of the linked SNP groups (Fig. S1). These selected SNPs were previously identified (D'Andrea *et al.* 2008); nevertheless, association studies have never been reported before for any of them.

Two pyrosequencing protocols (Table S2) were implemented to genotype the selected SNPs in 601 F_2 and F_3 pigs, one of which allowed the simultaneous genotyping of *LEP* g.1382C>T and *LEP* g.1387C>T. The genotype determinations were carried out using a PSQ-HS96 device (Pyrosequencing AB). Most of the *LEPR* c.1987C>T genotypes analysed in the current study were previously obtained in the study of Óvilo *et al.* (2005), but 110 new genotypes were obtained following the protocol by Óvilo *et al.* (2005). The *LEP g.1382C>T* and *LEP g.1387C>T* SNPs appeared fully linked in the genotyped animals, showing intermediate allele frequencies [(minor allele frequency MAF) = 0.41]. The results are reported in reference to the *LEP g.1387C>T* SNP, as representative of both SNPs. The *LEP g.1723A>G* SNP showed more extreme frequencies (MAF = 0.12).

Associations of *LEPR c.1987C>T* and/or *LEP g.1387C>T* SNPs with different traits were estimated using an animal model accounting for gender (male or female), batch (nine levels) and diverse covariates (age, carcass weight or backfat thickness). Likelihood ratio tests were calculated comparing the reduced and full models. The *P*-value for each additive or dominant effect was calculated comparing the full model with a model without the corresponding effect, assuming a chi-square distribution of the likelihood ratio with one degree of freedom. All statistical analyses were carried out with QXPAK version 3.0 software (Pérez-Enciso & Misztal 2004).

The individual analyses testing the effects of each gene revealed similar results to the joint analyses (Tables S4 and S5). However, given the known effects of the *LEPR* polymorphism on the analysed traits, the more reliable model to analyse *LEP* polymorphisms effects would be the one fitting the *LEPR* effects. Therefore, the main results of the association analyses discussed below correspond to the model fitting both genes. The *LEP g.1723A>G* SNP did not show any significant association. Nevertheless, the *LEP g.1387C>T* analysis revealed additive effects on the three growth-related traits, with the effect on body weight at 175 days of mean age (W175d) being statistically significant (Table 1). The *LEP* g.1387T allele led to an increase of around 1.5 kg body weight. The results obtained for the *LEPR* c.1987C>T polymorphism revealed, as expected, highly significant effects, mainly additive, on all the growth traits. The magnitudes of the effects of the *T* allele were larger than 2 kg. As the significant effects on carcass weight (CW) were additive for both *LEP* and *LEPR* polymorphisms, an additional analysis was conducted excluding the dominant effects from the model. The results revealed larger and more significant effects, essentially for the *LEP* polymorphism (1.227 ± 0.541; $P < 0.05$).

Significant associations of the *LEP* g.1387C>T polymorphism were also found with most of the fatness traits, such as backfat thickness at first rib (BFT1), backfat thickness between third and fourth ribs (BFT3) and belly bacon weight (BBW), with mainly dominant effects (Table 1). Although these dominant effects are not easy to explain, they are consistent and in the same direction across the measures, supporting their reliability. Regarding the *LEPR* c.1987C>T polymorphism, the analyses showed highly significant effects on all the analysed fatness traits. The effects were mainly additive, although the dominant effects also reached statistical significance (Table 1). In addition, significant effects were found for conformation traits. The *LEP* g.1387C>T polymorphism was associated with bone-in loin weight (BLW) (Table 1). The *LEPR* c.1987C>T polymorphism analysis revealed significant associations with all the analysed conformation traits except with HW (Table 1).

Associations of the *LEP* and *LEPR* polymorphisms with fatty acid composition in this animal material are reported for the first time in the present study (Table 2). The *LEP g.1387C>T* SNP showed significant additive effects on C18:1(n-9) and C18:2 fatty acids and the MUFA index, and dominant effects on C14:0, C16:0, C16:1, C18:2 and C20:2 fatty acids and the PUFA index. Significant effects of the *LEPR* c.1987C>T polymorphism were detected on C14:0, C16:0, C18:0, C18:1 (n-9), C18:2 and C20:1 fatty acids and the SFA and PUFA indexes; these were mainly additive effects, except for C20:1. The effects of the *LEPR*

Table 1 Results of the joint *LEPR* c.1987C>T and *LEP* g.1387C>T polymorphism analysis with growth, fatness and conformation traits.

	LEP g.1387T				*LEPR* c.1987T			
Trait	Additive (SE)	*P*-value	Dominant (SE)	*P*-value	Additive (SE)	*P*-value	Dominant (SE)	*P*-value
Growth-related traits								
W150d	1.42 (0.78)	0.07	0.36 (1.03)	0.73	2.67 (0.77)	6.1×10^{-4}	−1.68 (0.99)	0.09
W175d	1.62 (0.69)	0.02	−0.21 (0.89)	0.82	2.86 (0.65)	1.3×10^{-5}	−2.01 (0.85)	0.02
CW	1.07 (0.56)	0.07	−0.80 (0.73)	0.27	2.14 (0.52)	4.3×10^{-5}	−0.85 (0.70)	0.22
Fatness-related traits								
BFT1	−0.01 (0.04)	0.81	0.11 (0.05)	0.03	0.26 (0.04)	2.0×10^{-11}	−0.05 (0.05)	0.34
BFT2	0.03 (0.03)	0.27	0.04 (0.04)	0.25	0.23 (0.03)	1.1×10^{-16}	−0.08 (0.04)	0.03
BFT3	0.05 (0.04)	0.25	0.15 (0.06)	0.01	0.37 (0.04)	1.0×10^{-16}	−0.13 (0.05)	0.02
BBW	0.06 (0.03)	0.07	0.15 (0.04)	4.1×10^{-4}	0.24 (0.03)	7.7×10^{-14}	−0.10 (0.04)	0.02
IMF	0.03 (0.04)	0.47	0.03 (0.05)	0.49	0.19 (0.04)	1.3×10^{-7}	−0.09 (0.05)	0.06
Conformation-related traits								
HW	0.04 (0.05)	0.42	0.02 (0.06)	0.70	0.03 (0.04)	0.43	−0.04 (0.06)	0.52
SW	0.01 (0.02)	0.79	−0.01 (0.03)	0.79	−0.09 (0.02)	2.0×10^{-4}	0.08 (0.03)	7.0×10^{-3}
BLW	0.15 (0.07)	0.03	0.17 (0.09)	0.06	0.20 (0.07)	2.0×10^{-3}	−0.10 (0.09)	0.28
EMA	−0.50 (0.29)	0.09	−0.68 (0.38)	0.07	−1.26 (0.28)	1.2×10^{-5}	1.13 (0.37)	3.0×10^{-3}

W150d: Body weight at 150 days of mean age; W175d: Body weight at 175 days of mean age; CW: Carcass weight; BFT1: Backfat thickness at first rib; BFT2: Backfat thickness at last rib; BFT3: Backfat thickness between third and fourth ribs; IMF: Intramuscular fat content in longissimus dorsi muscle; BBW: Belly bacon weight; EMA: Longissimus dorsi eye muscle area; HW: Ham weight; SW: Shoulder weight; BLW: Bone-in loin weight.

c.1987C>T SNP were consistent with the effects reported in intramuscular and subcutaneous fat composition in previous studies (Muñoz *et al.* 2011; Rodríguez *et al.* 2010; Galve *et al.* submitted). However, the *LEP* polymorphisms had not previously been associated with fatty acid composition in pigs. An additional association analysis was performed using backfat thickness as a covariate in the model, instead of carcass weight. The analyses showed that the statistical significance and magnitude of the *LEPR* and *LEP* effects decreased and even disappeared for several fatty acids. The content of some fatty acids is highly correlated with backfat depth (Clop *et al.* 2003), and therefore, the effects of *LEPR* and *LEP* polymorphisms detected on fatty acid composition could be mediated through the effects on fatness. The *T* alleles of both *LEP* and *LEPR* SNPs would lead to an increase in voluntary feed intake, fatness and thus SFA content and to a decrease in MUFA and PUFA fatty acids. Conversely, the *LEP* g.1387T allele showed an opposite effect to the one expected for the MUFA index (Table 2). Moreover, this effect did not disappear when including BFT as a covariate in the model. This unexpected effect could be explained by a local effect of LEP on the oleic fatty acid level (William *et al.* 2002).

Additional analyses were performed accounting for additive × additive, additive × dominance, dominance × additive and dominance × dominance interaction effects between both SNPs, following Cockerham's model for epistatic interactions (Cockerham 1954). These analyses did

Table 2 Results of the joint *LEPR* c.1987C>T and *LEP* g.1387C>T polymorphisms analysis with fatty acid composition traits in subcutaneous fat.

	LEP g.1387T				*LEPR* c.1987T			
Trait	Additive (SE)	*P*-value	Dominant (SE)	*P*-value	Additive (SE)	*P*-value	Dominant (SE)	*P*-value
Fatty acid percentages in subcutaneous fat								
C14:0	−0.01 (0.01)	0.66	0.04 (0.02)	5.2×10^{-5}	0.03 (0.01)	1.6×10^{-5}	−0.01 (0.01)	0.91
C16:0	−0.13 (0.08)	0.10	0.26 (0.10)	0.01	0.30 (0.08)	9.3×10^{-5}	−0.17 (0.99)	0.09
C16:1	−0.03 (0.02)	0.22	0.07 (0.03)	0.02	0.03 (0.02)	0.26	0.02 (0.03)	0.45
C18:0	0.06 (0.08)	0.49	−0.06 (0.11)	0.60	0.18 (0.08)	0.03	−0.16 (0.11)	0.14
C18:1 (n-9)	0.28 (0.11)	0.01	0.12 (0.14)	0.40	−0.21 (0.10)	0.05	0.02 (0.14)	0.87
C18:1 (n-7)	−0.01 (0.02)	0.94	0.04 (0.03)	0.20	−0.02 (0.02)	0.37	0.02 (0.03)	0.48
C18:2	−0.17 (0.09)	0.05	−0.41 (0.12)	5.4×10^{-4}	−0.28 (0.09)	1.0×10^{-3}	0.29 (0.11)	0.01
C18:3	−0.02 (0.01)	0.21	−0.01 (0.02)	0.38	−0.02 (0.01)	0.11	0.02 (0.02)	0.10
C20:1	−0.01 (0.01)	0.64	−0.01 (0.02)	0.65	−0.01 (0.01)	0.63	−0.05 (0.02)	4.1×10^{-3}
C20:2	−0.01 (0.01)	0.45	−0.05 (0.01)	8.3×10^{-4}	−0.01 (0.01)	0.38	−0.01 (0.01)	0.75
SFA	−0.07 (0.12)	0.56	0.24 (0.16)	0.14	0.51 (0.12)	2.0×10^{-5}	−0.33 (0.16)	0.04
MUFA	0.26 (0.13)	0.04	0.23 (0.17)	0.18	−0.21 (0.12)	0.09	0.02 (0.16)	0.89
PUFA	−0.19 (0.10)	0.05	−0.46 (0.13)	2.8×10^{-4}	−0.30 (0.09)	1.0×10^{-3}	0.31 (0.12)	0.01

SFA: Saturated fatty acid sum; MUFA: Monounsaturated fatty acid sum; PUFA: Polyunsaturated fatty acid sum.

not reveal any significant effect of the interaction on growth, fatness and fatty acid composition. Significant effects of the *LEPR* c.1987C>T and *LEP* g.1387C>T interaction were detected for shoulder weight (SW) and BLW traits. For SW, the interactions were additive × additive (-0.081 ± 0.032; $P < 0.05$) and additive × dominant (0.139 ± 0.043; $P < 0.01$), and for BLW, it was additive × additive (-0.191 ± 0.093; $P < 0.05$). However, the low proportion of interactions found, two of 100, lies below the expected proportion of false discoveries. Therefore, the detected interaction effects are probably due to chance.

The effects of the *LEPR* c.1987C>T polymorphism are more consistent than the *LEP* g.1387C>T ones, probably because it is unlikely that *LEP* g.1387C>T or *LEP* g.1382C>T intronic polymorphisms would be the actual causal mutation, although they should be in linkage disequilibrium with it, while the *LEPR* c.1987C>T polymorphism seems to be potentially causal. A more stringent analysis would require the adjustment for multiple testing of the 25 highly correlated traits. Nevertheless, these results constitute new evidence for the relevance of *LEP* polymorphisms, although the potential usefulness of the quoted *LEP* SNPs would require the validation of the effects detected here in other pig populations.

Acknowledgements

This work was funded by MICINN project AGL2008-04818-C03/GAN. Dafne Pérez-Montarelo was funded by an FPI PhD grant from the Spanish Ministerio de Educación (BES-2009-025417). We thank Angelines López, Rita Benítez and Yolanda Nuñez for their technical assistance.

References

Amills M., Villalba D., Tor M., Mercadé A., Gallardo D., Cabrera B., Jiménez N., Noguera J.L., Sánchez A. & Estany J. (2008) Plasma leptin levels in pigs with different *leptin* and *leptin receptor* genotypes. *Journal of Animal Breeding and Genetics* **15**, 228–33.

Barb C.R., Hausman G.J. & Houseknecht K.L. (2001) Biology of leptin in the pig. *Domestic Animal Endocrinology* **21**, 297–317.

Chen C.C., Chang T. & Su H.Y. (2004) Genetic polymorphisms in porcine *leptin* gene and their association with reproduction and production traits. *Australian Journal of Agricultural Research* **55**, 699–704.

Clop A., Ovilo C., Perez-Enciso M. *et al.* (2003) Detection of QTL affecting fatty acid composition in the pig. *Mammalian Genome* **14**, 650–6.

Cockerham C.C. (1954) An extension of the concept of partitioning ward to extend Cockerham's model to consider multiple hereditary variance for analysis of covariances among relatives loci without further setting the orthogonal contrast when epistasis is present. *Genetics* **39**, 859–82.

D'Andrea M., Pilla F., Giuffra E., Waddington D. & Archibald A.L. (2008) Structural analysis and haplotype diversity in swine *LEP* and *MC4R* genes. *Journal of Animal Breeding and Genetics* **125**, 130–6.

De Oliveira Peixoto J., Facioni S.E., Savio P., Menck M.A., Vieira A., Gualberto M.V., De Almeida R., De Almeida E. & Silva M. (2006) Associations of *leptin* gene polymorphisms with production traits in pigs. *Journal of Animal Breeding and Genetics* **123**, 378–83.

Galve A., Burgos C., Silió L., Varona L., Rodríguez M.C., Óvilo C., Lopez-Buesa P. (In Press) The effects of leptin receptor (*LEPR*) and melanocortin-4 receptor (MC4R) polymorphisms in fat content, fat distribution, and fat composition in a Duroc × Landrace/large White cross. Livestock Science, Doi: 10.1016/j.livsci.2012.01.010.

Houseknecht K.L., Baile C.A., Matteri R.L. & Spurlock M.E. (1998) The biology of leptin: a review. *Journal of Animal Science* **76**, 1405–20.

Jiang Z.H. & Gibson J.P. (1999) Genetic polymorphisms in the *leptin* gene and their association with fatness in four pig breeds. *Mammalian Genome* **10**, 191–3.

Kennes Y.M., Murphy B.D., Pothier F. & Palin M.F. (2001) Characterization of swine *leptin* (*LEP*) polymorphisms and their association with production traits. *Animal Genetics* **32**, 215–8.

Muñoz G., Óvilo C., Silió L., Tomás A., Noguera J.L. & Rodriguez M.C. (2009) Single- and joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on *Sus scrofa* chromosome 6 and leptin receptor effects on fatness and growth traits. *Journal of Animal Science* **87**, 459–68.

Muñoz G., Alcázar E., Fernández A., Barragán C., Carrasco A., de Pedro E., Silió L., Sánchez J.L. & Rodríguez M.C. (2011) Effects of porcine *MC4R* and *LEPR* polymorphisms, gender and Duroc sire line on economic traits in Duroc × Iberian crossbred pigs. *Meat Science* **88**, 169–73.

Óvilo C., Perez-Enciso M., Barragan C., Clop A., Rodriguez C., Oliver M.A., Toro M.A. & Noguera J.L. (2000) A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6. *Mammalian Genome* **11**, 344–6.

Óvilo C., Oliver A., Noguera J.L. *et al.* (2002) Test for positional candidate genes for body composition on pig chromosome 6. *Genetic Selection and Evolution* **34**, 465–79.

Óvilo C., Fernández A., Noguera J.L. *et al.* (2005) Fine mapping of porcine chromosome 6 QTL and *LEPR* effects on body composition in multiple generations of an Iberian by Landrace intercross. *Genetic Research Cambridge* **88**, 57–67.

Óvilo C., Fernández A., Fernández A.I., Folch J.M., Varona L., Benítez R., Nuñez Y., Rodríguez C. & Silió L. (2010) Hypothalamic expression of porcine *leptin receptor* (*LEPR*), *neuropeptide Y* (*NPY*), and cocaine- and amphetamine-regulated transcript (CART) genes is influenced by *LEPR* genotype. *Mammalian Genome* **21**, 583–91.

Pérez-Enciso M. & Misztal I. (2004) QXPAK: a versatile mixed model application for genetical genomics and QTL analyses. *Bioinformatics* **20**, 2792–8.

Rodríguez M.C., Fernández A., Carrasco C., García A., Gómez E., de Mercado E., López M.A., Óvilo C. & Silió L. (2010) Effect of *LEPR* c.2002C>T SNP on feed intake and growth in heavy Duroc × Iberian crossbred pigs. In *Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production*: 1–6 August 2010; Leipzig.

Switonski M., Stachowiak M., Cieslak J., Bartz M. & Grzes M. (2010) Genetics of fat tissue accumulation in pigs: a comparative approach. *Journal of Applied Genetics* **51**, 153–68.

William W.N. Jr, Ceddia R.B. & Curi R. (2002) Leptin controls the fate of fatty acids in isolated rat white adipocytes. *Journal of Endocrinology* **175**, 735–44.

Wylie A.R.G. (2011) Leptin in farm animals: where are we and where can we go? *Animal* **5**, 246–67.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 Scheme of the porcine *LEP* gene fragment sequenced according to the reference sequence (GenBank AJ865080.1).

Table S1 Phenotypic traits registered in F_2 and F_3 generations of the Iberian × Landrace cross.

Table S2 Primer pairs used for *LEP* sequencing and polymorphisms genotyping.

Table S3 Description of the novel polymorphisms identified in *LEP* gene by sequencing the Iberian and Landrace parental pigs.

Table S4 Results of *LEP* g.1387C>T polymorphism analysis with growth, fatness and conformation traits.

Table S5 Results of *LEPR* c.1987C>T polymorphism analysis with growth, fatness and conformation traits.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

ARTÍCULO IV

Haplotypic diversity of porcine *LEP* and *LEPR* genes involved in growth and fatness regulation

Haplotypic diversity of porcine *LEP* and *LEPR* genes involved in growth and fatness regulation

Dafne Pérez-Montarelo[1*], M. Carmen Rodríguez[1], Rita Benítez[1], Fabián García[1], Luis Silió[1] and Ana I. Fernández[1].

[1]Mejora Genética Animal, Instituto Nacional de Tecnología Agraria y Alimentaria, 28040 Madrid, Spain.

*Corresponding autor: dafne.perez@inia.es. Phone: (0034)913473792.

ABSTRACT

The analysis of the structural genetic variability in candidate genes can make it possible to analyze the selection footprint and deepen the understanding of the genetic basis of complex traits. The leptin (*LEP*) and its receptor (*LEPR*) genes are involved in food intake and energy homeostasis and polymorphisms associated to growth and fatness traits have been previously detected in both genes. The main objective of this study was to explore the genetic variability of the most polymorphic regions of both *LEP* and *LEPR* genes in a variety of pig populations and wild boars from European and Asiatic origins. In total, 46 animals were included in the analyses: 12 Iberian pigs, 20 pigs from European domestic breeds, 9 Spanish wild boars and 5 pigs from Asiatic domestic breeds. The sequencing of a total of 2,520 bp of the intron 2-3 of the *LEP* gene and 1,266 bp of the promoter region of the *LEPR* gene, allowed the identification of 68 and 24 polymorphisms for *LEP* and *LEPR* genes, respectively. The nucleotidic and haplotypic variation was studied in both genes. Neighbour-joining trees and Median-joining networks were built for the 68 haplotypes identified in the *LEP* gene and the 16 haplotypes detected in the *LEPR*, using the Phase2.1 and Network software. The results of both genes showed the known genetic divergence between European and Asiatic pig breeds. An extremely high variability of the *LEP* gene was detected in the different analyzed populations and allowed the confirmation of the existence of two domestication centers in Asia. In comparison to the *LEP* gene, the *LEPR* showed a lower variability, especially in the Iberian breed that showed no variability supporting a possible selection event of the *LEPR* gene in this breed related with its high appetite and leptin levels that should be further investigated.

INTRODUCTION

The variability of the pig genome has been shaped by diverse forces such as multiple domestication events, populations admixture, natural selection and selective breeding Since the beginning of the 1990's, molecular data have played an essential role for characterizing the genetic diversity of farm animals at level of population, breed and

species (Frankham et al. 2002). Most of the studies performed on pigs have been conducted using specific mitochondrial DNA (mtDNA) regions (Giuffra et al. 2000; Alves et al. 2003; Kim et al. 2005; Luetkemeier et al. 2010) or microsatellites markers. (Laval et al. 2000; Martínez et al. 2000; Lemus-Flores et al. 2001; SanCristobal et al. 2006; Thuy et al. 2006; Sollero et al. 2009; Luetkemeier et al. 2010). Recently, a phylogenomic analysis of complete genome sequences of wild boars and Asiatic and Western domestic pigs substantiate the hypothesis that pigs were independently domesticated in Eurasia and East Asia and revealed Asian influence in most of the cosmopolitan European and American breeds (Groenen et al. 2013). Even more, the new massive parallel sequencing technologies have allowed the identification of some candidate regions within the porcine genome that putatively have been under selection for diverse goals (Amaral et al. 2011; Wilkinson et al. 2013).

Moreover, the analysis of the structural genetic variability in particular candidate genes for economically important traits can deepen the understanding of the genetic basis of such complex traits (D'Andrea et al. 2008; Yang et al. 2012). Because modern selective breeding towards leaner pigs must have dramatically affected the regulation of biological pathways related to growth and body composition; it will be particularly interesting to know the selection footprint left on genes related to these traits (Ojeda et al. 2006). There have been previous studies aimed at analyzing the genetic diversity of singular genes associated to growth and fatness traits such as *IGF2* (Ojeda et al. 2006), *FABP4* (Ojeda et al. 2008), *SERPINA*6 (Esteve et al. 2011), *PPARD* (Ren et al. 2011), *FTO* (Fontanesi and Russo 2012) and *MUC4* (Yang et al. 2012). The objective of the present study was to analyze the haplotypic variability of two key genes implicated in growth and fatness regulation, the leptin (*LEP*) and its receptor (*LEPR*), in different pig populations. The leptin and its receptor are involved in the regulation of food intake and energy balance (Friedman 2002). These genes have been widely studied in pigs due to their relevance on important economic traits (Wylie

2011, Switonski et al. 2010). Our previous studies on an Iberian x Landrace experimental cross reported significant effects of SNPs located in both genes on pig productive traits (Ovilo et al. 2010; Pérez-Montarelo et al. 2012 and 2013). The study of these genes results especially interesting in the Iberian breed, due to their particular obese phenotype and because they carry a fixed *LEPR*c1987T allele that reduces leptin signaling, enhances feed intake and increases fat content (Ovilo et al. 2010). The analysis of the *LEP* and *LEPR* gene sequences allowed the identification of their most polymorphic regions, which correspond to the promoter region of the *LEPR* gene (Pérez-Montarelo et al. 2013) and the intron between exons two and three of the *LEP* gene (Pérez-Montarelo et al. 2012). In the present study, the analyses of these *LEP* and *LEPR* polymorphic regions have been conducted in several pig populations from different origins to explore the haplotypic diversity and to obtain further insights into the breed-specific role of both genes on the control of voluntary feed intake and related productive traits.

MATERIAL AND METHODS

Pig breeds and specimens

A total of 46 animals were included in the analyses (Table 1). It comprises 28 domestic pigs from nine different Western breeds including 12 Iberian pigs from different varieties and strains representative of the breed (Torbiscal, Entrepelado, Retinto and Black Hairless) and 18 other European pig breeds: American Duroc (*n*=6), ancient Duroc Jersey (*n*=2), European Landrace (n=2), Large White (*n*=2), ancient Large White (*n*=2) preserved in Spain from 1931, Pietrain (*n*=2), and the endangered breeds UK Large Black (*n*=1) and French Noir de Bigorre (*n*=1). Besides, nine European wild boars from different Spanish regions were also included in the analyses. In addition, two Chinese Meishan domestic pigs, three Vietnamese pigs and two pigs of a commercial cross (Youli) between Landrace and a synthetic Chinese-European line were also included in the analyses.

DNA extraction and sequencing

Genomic DNA from all animals was extracted from blood samples with a standard phenol: chloroform protocol

(Sambrook et al. 1989), and used for sequencing and polymorphisms identification. Four primer pairs were designed according to the reference sequence GenBank: AJ865080.1 of the porcine *LEP* gene to amplify 2,520 bp of the intronic region between exons two and three (Supplementary table 1). The *LEPR* promoter region sequencing was defined in accordance with the pig *LEPR* gene structure described by Lee et al. (2008). According to the available *LEPR* gene sequence (GenBank: FN677933.1), three primer pairs (Supplementary table 1) were designed to amplify 1,266 bp, in three overlapped fragments.

PCRs were carried out in a 25µl final volume containing 100ng of DNA, 1 unit of polymerase (Biotools) or HotStart polymerase (Quiagen), specific buffer, 2mM of dNTPs and 0.5 µM of each primer. The specific annealing temperature of each primer pair is shown in Supplementary table 1. The PCR reactions were carried out in a GeneAmp PCR System 9700 (Applied Biosystems, Warrington, UK). The PCR products were purified with the GFX™ PCR DNA purification kit (GE Healthcare, UK) according to the manufacturers' protocol. All products were sequenced using the 3100 BigDye® Terminator v3.1 Matrix Standard in a 3730 DNA Analyzer (Applied Biosystems Warrington, UK). The obtained sequences were edited and aligned using the EditSeq and MegAlign packages of the WinStar software for the identification of polymorphisms.

Data analysis

Sequence overhangs were trimmed resulting in a total of 2,480 bp aligned region for the *LEP* gene and 1,200 bp for the *LEPR*. The different haplotypes of both genes were independently constructed with Phase v2.1.1 (Li and Stephens 2003) using default options.

Multiple alignments of all sequences per gene was performed with the Molecular Evolutionary Genetics Analysis version 5.05 (MEGA5) software (Edgar 2004; Tamura et al. 2011) and haplotypic dendrograms were inferred using the Neighbour-Joining model using the pair-wise distances (p-distance). To assess the robustness of the dendrogram topology, bootstrap resampling was carried out with 1000 replicates. Between and within

haplotypes groups and net p-distances were calculated using the MEGA5 software. The nucleotide variability of each one of the groups of pig populations was estimated calculating the number (S) and the proportion (Ps) of segregating sites and the nucleotide diversity index (π). The haplotipic variation was also analyzed by the number of haplotypes (h) and the allelic richness (ph) or expected number of haplotypes calculated using the rarefaction method described by El Mousadik and Petit (1995). The visualization of the pair-wise distance values calculated with MEGA5 between each pair of haplotypes as a heatmap was performed with the Multiple Experiment Viewer (MeV) software of the TM4 Microarray Software Suite (Saceed et al. 2006) for genomic data exploration. In addition, the Median Joining network-building option of the Network v4.6.1.1 software (Bandelt et al. 1999) was used using the default settings to construct phylogenetic networks of the identified haplotypes. This software uses the number of nucleotide substitutions as a measure of the distance between the haplotypes to build the network. The MP option (Polzin et al. 2003) was used to clean the network by eliminating unnecessary median vectors and links.

LEPR gene expression conditional on promoter haplotype

The expression level of the *LEPR* gene (long isoform, *LEPRb*) was analyzed in 30 pigs from an F1 (Iberian x Landrace) x Landrace backcross using the protocol described by Óvilo et al. (2005). The *LEPR* promoter region was also sequenced in this 30 animals and their haplotypic genotypes were identified as mentioned before. The differential expression of this gene according to the haplotypes was investigated comparing the expression in those animals carrying haplotype *LEPR*H1 with the ones carrying other haplotypes. Expression was quantified using the method proposed by Steibel et al. (2009).

RESULTS

Nucleotidic variation

The alignment of the sequences of the *LEP* gene intron 2-3 revealed a total of 68 polymorphisms in the 46 sequenced pigs. Three of these polymorphisms correspond to indels

and the remaining 65 to SNPs (7 of which were singletons). The number and the proportion of segregating sites and the nucleotide diversity for the different groups of pig breeds or populations (Spanish wild boars, Iberian domestic breed, other European breeds and Asiatic breeds) are shown in Table 1. The highest nucleotidic variation for this gene was detected in the Asiatic domestic breeds, with 55 segregating polymorphisms out of the 68 detected, followed by the European domestic breeds (excluding the Iberian breed) that showed 41segregating sites. Similarly, the highest nucleotide diversity index (π) was found in the Asiatic domestic breeds, with a value of 0.38, followed by the European domestic breeds (excluding the Iberian breed) with an index of 0.24. The Iberian pigs and the Spanish wild boars showed the lowest nucleotide diversity index, being very similar in both cases, around 0.19. Moreover, the nucleotidic variation of these two groups was also similar with 36 segregating polymorphisms in Iberian pigs and 37 in Spanish wild boars.

The alignment of the sequences of the promoter region of *LEPR* gene revealed 24 polymorphisms, two indels and 22 SNPs (0 singletons). The number and the proportion of segregating sites and the nucleotide diversity for the different groups of populations are shown in Table 1. The European domestic breeds (excluding the Iberian breed) showed the highest nucleotidic variation, with 20 segregating polymorphisms out of the 24 detected, followed by wild boars that showed 18 segregating sites. Similar nucleotide diversity indexes, around 0.004, were found in the different populations, except in the Iberian breed. It is remarkable the fact that no nucleotidic variation was found in the promoter region of the *LEPR* gene in the analyzed Iberian pigs.

Haplotypic diversity

A total number of 68 segregating haplotypes were constructed for the *LEP* gene, most of them were found at low frequencies (ranging from 1/56 to 6/56). Based on the 68 haplotypes identified for *LEP* gene, a dendrogram was constructed (Figure 1). Four major clades are distinguishable, named from A to D. The clades A and

B of the dendrogram comprise European pig breeds, including Iberian and wild boar haplotypes. The clades C and D contain the 12 haplotypes of likely Asiatic origin, clade C is composed by the five haplotypes detected only in Vietnamese Asiatic pig breed (*LEP*H7, H23, H24, H42 and H43), and clade D contains the four haplotypes detected in the Meishan animals (*LEP*H32, H61, H63 and H64), one from Youli (*LEP*H62) and two haplotypes detected in Duroc pigs (*LEP*H65 and H66). The dendrogram also reflects the genetic distances between and within groups. The average within, between and net between groups genetic distances for the different groups (European 1 and 2, Chinese and Vietnamese) or clades (A, B, C and D) are shown in Table 2a. It is remarkable that in all cases the average within-groups distances are smaller than the between groups distances. The Chinese clade (D) showed the highest within group distance and the Vietnamese (C) the lowest. The highest between group distances were found between Chinese and Vietnamese clades and the lowest between the two European clades (A and B). In agreement with this result, the heatmap obtained for this gene (Figure 2), indicating the distances between each pair of haplotypes, shows that the largest distances are those obtained between the haplotypes of the Asiatic subclusters (*LEP*H7, H23, H24, H32, H42, H43, H61, H62, H63, H64, H65, H66) and the rest, while the shortest distances are found among the four blocks of European haplotypes, three in subcluster A (one from LEP3 to LEP6, another from LEP15 to LEP22 and the last one from LEP35 to LEP41) and one in subcluster B (from LEP48 to LEP60). The same four subclusters can be found in the MJ network of *LEP* haplotypes and they have been named in the same way (Figure 3). It is remarkable that the two subclusters from Asiatic origin (C and D) appeared completely separated in both extremes of the network. Regarding the European subclusters, the A subcluster contains a high representation of haplotypes from Iberian pigs and Spanish wild boars, although other European domestic breeds are also present. Conversely, subcluster B is mainly formed by other European domestic pig breeds excluding Iberian, even though some Iberian and Spanish

wild boars are present also in this subcluster. Moreover, some of the European haplotypes are shared between Iberian pigs and wild boars (*LEP*H16 and *LEP*H34), between Iberian pigs and other European domestic breeds (*LEP*H5 and *LEP*H51) and between wild boars and the European domestic breeds, excluding Iberian (*LEP*H52 and *LEP*H60). It is also worth mentioning that the wild boar haplotype *LEP*H13 is linked to European subcluster A but far away in the network.

The haplotype number (h), the haplotypic richness (hr) and the number private haplotypes (ph) for the different populations are shown in Table 1. The highest number of haplotypes was found in the group of other European domestic breeds, excluding Iberian, and the lowest was detected in the Asiatic domestic breeds. The haplotypic richness calculated using the correction by the number of haplotypes per population through rarefaction, values adjusted to the smallest group size, ten, allowed us to confirm a higher haplotypic variation in the other European domestic breed group, than in Asiatic domestic breed group. The Spanish wild boars have the same haplotypic variation than other European domestic breeds. In addition, the correction by rarefaction performed allowed us to determine that the haplotypic variation was higher in wild boars than in Iberian pigs. The results evidenced a high haplotypic variation in this gene, especially among European domestic breeds and wild boars, but even within the Iberian breed that showed a total of 18 different haplotypes in the 12 individuals sequenced.

A total of 16 segregating haplotypes were reconstructed for the *LEPR* gene. Among the 16 identified *LEPR* haplotypes, the *LEPR*H1 showed the highest frequency (equal to 51/92) while the others ranged from 1/92 to 8/92. The constructed dendrogram based on the 16 haplotypes identified for *LEPR* gene (Figure 4) shows two main clades. The upper major clade of the dendrogram comprises European pig breeds including Iberian and the European wild boars haplotypes (*LEPR*H1, H2, H3, H4, H5, H6, H11 and H13). The lower clade contains all the Asiatic pig breeds haplotypes (*LEPR*H7 and H16 detected in Chinese pigs and *LEPR*H8 identified in

Vietnamese pigs.) together with some haplotypes detected in European breeds (*LEPR*H8, H9, H10, H14, H15 and H16 detected in Large Black, Large White, Youli, Landrace and Duroc breeds). Finally, while most of the wild boar haplotypes are in the European clade, there is one wild boar haplotype (*LEPR*H12) that appeared separated between the quoted clades. The average within, between and net between groups distances are shown in Table2b. The within-group genetic distance were similar for both European and Asiatic groups and lower than the between and net between group mean distances. The *LEPR* heatmap (Figure 5) evidences that the longest distances are obtained within some of the European and Asiatic haplotypes (*LEPR*H11 and H13 and the rest of the European haplotypes, and *LEPR*H10 and H14 and the rest of the Asian haplotypes). The MJ network constructed for the *LEPR* gene haplotypes (Figure 6) shows the same two subclusters, A subcluster with haplotypes detected in both Asiatic and European pig breeds and subcluster formed by haplotypes identified in both European pig breeds and wild boars. It could be seen that some haplotypes are shared between different origins, haplotypes *LEPR*H8 and *LEPR*H16 are present in both Asiatic and European pig breeds, and haplotypes *LEPR*H1 and *LEPR*H5 are present in European pig breeds and in wild boars. It is noteworthy that all 12 Iberian pigs have the same *LEPR*H1 haplotype. In contrast, seven different haplotypes were identified in the nine wild boars from different origins in the Iberian Peninsula included in this analysis. As it is shown in the network, five of these seven haplotypes were specifically identified in wild boars (*LEPR*H12, *LEPR*H3, *LEPR*H2, *LEPR*H11 and *LEPR*H6) and the remaining two are shared with European pig breeds (*LEPR*H1 and *LEPR*H5). They all cluster around the European pig breeds except haplotype *LEPR*H12.

The haplotypes number (h), the haplotypic richness (hr) and the number private haplotypes (ph) for the different populations are shown in Table 1. Even though the observed number of haplotypes indicates that the highest haplotypic variation is found in the other European breed group, excluding Iberian, the correction by rarefaction performed taking into account the different

sample size of the populations, allowed to determine that the highest haplotipic variation is indeed found in wild boars. In agreement with the nucleotidic variation found for this gene in the Iberian population, all the Iberian samples analyzed shared the same *LEPR*H1 haplotype, having no haplotypic variation and zero private haplotypes.

The analysis of the promoter region sequence in 30 animals of the Iberian x Landrace backcross allowed us to identify the *LEPR*H1 haplotype in 42 out of the 60 sequences available from the 30 animals and 18 carriers of four other haplotypes, which were not identified in the previous analysis. The comparison of the *LEPR*H1 carriers gene expression levels versus the remaining ones did not reveal significant results (*P-value* =0.39).

DISCUSSION

The most polymorphic regions of the *LEP* and *LEPR* genes have been sequenced in a variety of pig breeds and wild boars, from Western and Asiatic origins, in order to analyze their nucleotidic and haplotypic diversity. These genes are involved in growth, fatness and energy balance regulation and have a relevant impact on pig productive traits. The interest of analyzing *LEP* gene comes from its highly polymorphic nature, previously reported by D'Andrea et al. (2008), who considered the *LEP* gene a hot spot with an extensive amount of polymorphisms. The results obtained here from the sequence variability analysis of the *LEP* intronic region are in agreement with those obtained from specific polymorphic regions in the mtDNA, evidencing once more the genetic differences between Asiatic and European haplotypes. Previous studies indicated multiple independent domestications of the pig mainly in Asia and Europe (Larson et al. 2005 and 2007) and dated from roughly three hundred years ago the moment when Chinese pigs came to Europe gathering two suid lines divided for more than nine thousand years (White 2011). These millennia of diverging genetic, cultural and evolutionary pressures created remarkably different phenotypes of both pig origins.

Moreover, as previously reported by Wu et al. (2007), our *LEP* gene

result supports the hypothesis of the existence of two different domestication events in two Asia regions, one domestication centre would be the Mekong region and the other one would be the downstream region of the Yangtze River. In fact, our net between groups mean distance pointed out that the distance between Meishan and European (0.27±0.04) and between Vietnamese and European (0.36±0.05) are smaller than the one obtained between both Asiatic breeds (0.45±0.05). According to this hypothesis, the Meishan pigs analyzed in the present study could likely be derived from the Mekong region and the Vietnamese pigs from the downstream region of the Yangtze River. It has also been previously reported the real shift in the pig populations occurred at the end of the eighteen century when Chinese breeds arrived to Britain playing a key role in the transformation of the English breeds producing improved crossbreeds combining the larger size of the European pigs with the fatter body and faster early growth of the Asiatic (White 2011). The clustering observed in the different analyses of the present study is supported by the documented history of "old" breeds. Here, the European and Asiatic breeds appeared as different clades in the dendrogram and as separated clusters in the built networks, however some European haplotypes (*LEP*H62, H65 and H66) from Duroc and Youli animals are included in the Chinese cluster giving further evidence of the known introgression of Chinese genomes in the European breeds during the XVIII and XIX centuries (Porter 1993; White 2011). The dendrograms showed a typical clustering of the commercial lines around their respective breeds of reference (Ollivier et al. 2009). For example, the haplotypes of Youli pigs (*LEP*H31, H54 and H62) clustered both with European (*LEP*H31 and H54) and Asiatic (*LEP*H62) pig breeds, because they come from a cross between Landrace boars and hyperprolific sows of a synthetic line generated from Chinese and European breeds. It is remarkable the fact that the same pattern showed previously with mtDNA and microsatellites markers distributed along the whole genome or even with complete genome sequences, is obtained here with the information of

a relatively small gene region. One interesting result is the identification of two European subclusters in MJ network of *LEP* haplotypes, one containing a high representation of haplotypes from Iberian pigs and Spanish wild boars but also other European domestic breeds and the other one mainly containing other European domestic pig breeds but also some Iberian and Spanish wild boars. This subdivision in the European pig breeds (Iberian, wild boars and other European breeds) has not been identified in previous studies and could have several explanations, as sampling bias or recombination events across haplotypes, which should be further investigated. In wild boars 17 different *LEP* haplotypes were detected in the nine animals sequenced, 13 of them were specifically detected in wild boars (*LEP*H1, H4, H13, H18, H19, H20, H21, H28, H29, H39, H40, H41 and H59), two were shared with Iberian pigs (*LEP*H16 and H34) and the remaining two were shared with European pig breeds (*LEP*H52 and H60). However, these observations do not agree with previous studies that reported very low nucleotide diversity in European wild boar compared to domestic breeds for the *FABP4* gene (Ojeda et al. 2006) and mtDNA (Larson et al. 2005; Fang and Andersson 2006). It is also worthy mention the high haplotype diversity detected in the Iberian breed, in which a total of 18 haplotypes were identified in only 12 animals. Besides, the network and the heatmap constructed reinforce the high haplotype diversity of this gene.

The interest of analyzing the *LEPR* gene is not only its variability, but also the previous results obtained for the *LEPR*c.1987C>T effect on growth and fatness Óvilo et al., 2005; Muñoz et al., 2009; Muñoz et al., 2010; Uemoto et al., 2012; Galve et al., 2012). The results obtained for this gene also support both the genetic differences between European and Asiatic breeds and the gene introgression of Asian into the Europeans. The fact that the Large Black sample shares haplotypes with Vietnamese pigs (*LEPR*H8) is in agreement with the previously reported introgression of Asian origin into the Large Black population shown in the *MC1R* gene (Kijas et al. 1998). Again, the wild boars from different origins of the Iberian Peninsula included in this study

revealed the highest *LEPR* haplotypic diversity. A total of seven *LEPR* different haplotypes were identified in the nine wild boars. Moreover, while most wild boar haplotypes clustered with the European pig breeds haplotypes, as expected, there is one haplotype (*LEPR*H12) that clustered separated from them as a different clade. The distances between *LEPR*H12 haplotype and the haplotypes contained in European and Asiatic clades of the dendrogram are 0.309±0.077 and 0.259±0.075, respectively. The identification of particular wild boar haplotypes has been reported in a previous study (Giufrra et al. 2000), where wild boars from Italy clustered also as a separate clade in the dendrogram. Moreover, in our results, five of the seven haplotypes were specifically detected in the wild boars (*LEPR*H2, H3, H6, H11 and H12), and only two were shared with European pig breeds (*LEPR*H1 and H5).

It is remarkable the null *LEPR* variability observed in the Iberian breed, in which only one haplotype (*LEPR*H1) has been identified. This low haplotipic variability of the *LEPR* gene (1/16 haplotypes) contrasts with the results obtained for the *LEP* gene (17/68 haplotypes) and with previous studies that reported a high genetic variability of the Iberian breed in the *FABP4* (Ojeda et al. 2006) and *IGF2* genes (Ojeda et al. 2008) and using mtDNA (Alves et al. 2003) or microsatellites markers (Fabuel et al. 2004; Rodrigañez et al. 2008; Gama et al. 2013). A plausible explanation for the low genetic diversity of the *LEPR* gene in Iberian pigs could be that it would have undergone whether natural or artificial selection due to its impact. Following this idea, some mutations would have been favored over others reducing the variability and the classical hitchhiking genetic effect predicts that a selective sweep will reduce genetic variability around the selected target (Ojeda et al. 2008). Here, it should be noticed the specific phenotype of the Iberian breed for growth and fatness to properly evaluate the results. Iberian pigs present an extremely fat phenotype, high levels of voluntary feed intake, high serum levels of circulating leptin and high lipogenic potential. All these characteristics fit with a leptin resistance pattern in which leptin levels fail to reduce voluntary feed intake and obesity (Fernández-

Figares et al. 2007; Ovilo et al. 2010). In this sense, polymorphisms in the *LEPR* gene that would reduce its expression or its ability to transmit leptin signal could be responsible of the phenotype. Previous studies of a putative causative mutation of this gene *LEPR*c.1987C>T in Iberian crosses support this idea (Ovilo et al. 2010). Iberian pigs have fixed the *LEPRc*.1987T allele, associated to a low *LEPR* mRNA expression, and probably leading to a reduction in leptin signaling that is translated in higher feed intake and fatness (Ovilo et al. 2010). In order to evaluate the potential effect of the *LEPR*H1 promoter haplotype, also fixed in Iberian, on its gene expression, an additional analysis was conducted in an Iberian x Landrace backcross. However, the results did not revealed changes in gene expression conditional on promoter haplotypes.

The results of the present study evidence the high variability of the *LEP* gene in the different analyzed populations verify the genetic divergence between Asiatic and Western pig breeds, and confirm the existence of two domestication centers in Asia (Central China and Southeast Asia). In contrast, the obtained results prove the low genetic variability of the *LEPR* gene within and between populations. Even more, the Iberian pigs showed no variability in this gene, contrasting with its high phenotypic, productive and genetic variation identified in previous studies (Alves et al. 2003; Fabuel et al. 2004; Ojeda et al. 2006; Ojeda et al. 2008; Rodrigañez et al. 2008; Gama et al., 2013). This result supports a possible selection event of the *LEPR* gene in this breed potentially related with its high appetite, fatness and leptin levels and to the special characteristics of its traditional production system in La Dehesa, with a strong seasonal variation of the resources (López-Bote, 1998).

REFERENCES

Alves E, Fernández AI, Fernández-Rodríguez A, Pérez-Montarelo D, Benitez R, Ovilo C, Rodríguez C, Silió L (2009). Identification of mitochondrial marke rs for genetic traceability of European wild boars and Iberian and Duroc pigs. Animal **3**(9):1216-23.

Alves E, Ovilo C, Rodríguez MC, Silió L (2003) Mitochondrial DNA sequence
variation and phylogenetic relationsh ips among Iberian pigs and other domestic and wild pig populations. Anim Genet **34**(5):319-24.

Amaral AJ, Ferretti L, Megens HJ, Crooijmans RP, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MA (2011) Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS One **6**(4):e14782.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol **16**(1):37-48.

D'Andrea M, Pilla F, Giuffra E, Waddington D, Archibald AL (2008) Structural analysis and haplotype diversity in swine LEP and MC4R genes. J Anim Breed Genet **125**(2):130-6.

Edgar, R. C (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792-1797.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.)] endemic to Morocco. Theoretical and Applied Genetics **92**:7, pp 832-839.

Esteve A, Ojeda A, Huang LS, Folch JM, Pérez-Enciso M (2011) Nucleotide variability of the porcine SERPINA6 gene and the origin of a putative causal mutation associated with meat quality. Anim Genet **42**(3):235-41.

Fabuel E, Barragán C, Silió L, Rodriguez MC, Toro MA (2004) Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. Heredity **93**, 104-113.

Fang M, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. Proceedings Biological Sciences/the Royal Society **273**:1803– 1810.

Fernandez-Figares I, Lachica M, Nieto R, Rivera-Ferre MG, Aguilera JF (2007) Serum profile of metabolites and hormones in obese (Iberian) and lean (Landrace) growing gilts fed balanced or lysine deficient diets. Livestock Science **110**:73–81

Fontanesi L, Russo V (2013) Nucleotide variability and haplotype heterogeneity at the porcine fat mass and obesity-associated (FTO) gene. Anim Genet **44**(1):96-100.

Foulley JL, van Schriek MG, Alderson L, Amigues Y, Bagga M, Boscher MY, Brugmans B, Cardellino R, Davoli R, Delgado JV, Fimland E, Gandini GC, Glodek P, Groenen MA, Hammond K, Harlizius B, Heuven H, Joosten R, Martinez AM, Matassino D, Meyer JN, Peleman J, Ramos AM, Rattink AP, Russo V, Siggens KW, Vega-Pla JL, Ollivier L (2006) Genetic diversity analysis using lowly polymorphic dominant markers: the example of AFLP in pigs. J Hered **97**(3):244-52.

Frankham, RJ, Ballou, JD, Briscoe DA and MacInnes, KH (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.

Friedman JM (2002) The function of leptin in nutrition, weight, and physiology. Nutr Rev **60**:S1-14.

Galve A, Burgos C, Silió L, Varona L, Rodríguez C, Óvilo C, López-Buesa P (2012) The effects of leptin receptor (LEPR) and melanocortin-4 receptor (MC4R) polymorphisms on fat content, fat distribution and fat composition in a Durco x Landrace/Large White cross. Livestock Science **145**:145-152.

Gama LT, Martínez AM, Carolino I, Landi V, Delgado JV, Vicente AA, Vega-Pla JL, Cortés O, Sousa CO; BIOPIG Consortium (2013) Genetic structure, relationships and admixture with wild relatives in native pig breeds from Iberia and its islands. Genet Sel Evol **14**;45:18.

Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L (2000) The origin of he domestic pig: independent domest ication and subsequent introgression. Genetics **154**(4):1785-91.

Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LA,Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S,Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi SH, Chow W, Clark RC, Clee C, Crooijmans RP, Dawson HD, Dehais

P,De Sapio F, Dibbits B, Drou N, Du ZQ, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GJ, Fowler KE, Fredholm M, Fritz E, Gilbert JG, Giuffra E, Gorodkin J,Griffin DK, Harrow JL, Hayward A, Howe K, Hu ZL, Humphray SJ, Hunt T, Hornshøj H, Jeon JT, Jern P, Jones M, Jurka J, Kanamori H, Kapetanovic R, Kim J,Kim JH, Kim KW, Kim TH, Larson G, Lee K, Lee KT, Leggett R, Lewin HA, Li Y, Liu W, Loveland JE, Lu Y, Lunney JK, Ma J, Madsen O, Mann K, Matthews L,McLaren S, Morozumi T, Murtaugh MP, Narayan J, Nguyen DT, Ni P, Oh SJ, Onteru S, Panitz F, Park EW, Park HS, Pascal G, Paudel Y, Perez-Enciso M,Ramirez-Gonzalez R, Reecy JM, Rodriguez-Zas S, Rohrer GA, Rund L, Sang Y, Schachtschneider K, Schraiber JG, Schwartz J, Scobie L, Scott C, Searle S,Servin B, Southey BR, Sperber G, Stadler P, Sweedler JV, Tafer H, Thomsen B, Wali R, Wang J, Wang J, White S, Xu X, Yerle M, Zhang G, Zhang J, Zhang J,Zhao S, Rogers J, Churcher C, Schook LB (2012)
Analyses of pig genomes provide insig ht into porcine demography and evol ution. Nature **491**(7424):393-8.

Kijas JM, Wales R, Törnsten A, Chardon P, Moller M, Andersson L (1998) Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. Genetics **50**(3):1177-85.

Kim KI, Lee JH, Li K, Zhang YP, Lee SS, Gongora J, Moran C (2002) Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. Anim Genet **33**(1):19-25.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science **307**(5715):1618-21.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phylogeography of wild b oar reveals multiple centers of pig do mestication. Science **307**(5715):1618-21.

Laval G, Iannuccelli N, Legault C, Milan D, Groenen MA, Giuffra E, Andersson L, Nissen PH, Jørgensen CB, Beeckmann P, Geldermann H, Foulley JL, Chevalet C, Ollivier L (2000) Genetic diversity of eleven European pig breeds. Genet Sel Evol **32**(2):187-203.

Lee KT, Hwang H, Kang KS, Park EW, Kin JH, et al. (2008) Genomic structure of porcine leptin receptor gene (LEPR)/LEPR overlapping transcript gene (LEPROT) and SNP discovery, association study. Congress of the International Society of Animal Genetics 2008.

Lemus-Flores C, Ulloa-Arvizu R, Ramos-Kuri M, Estrada FJ, Alonso RA (2001) Genetic analysis of Mexican hairless pig populations. J ANim Sci **79**:3021-3026.

Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics **165**(4):2213-33.

López-Bote CJ (1998) Sustained utilization of the Iberian pig breed. Meat Science **49**S1:S17-27.

Luetkemeier ES, Sodhi M, Schook L B, Malhi RS
(2010) Multiple Asian pig origins revealed through genomic analyses. 2010. Mol Phylogenet Evol **54**:680-6.

Marincs F, Molnár J, Tóth G, Stéger V, Barta E (2013) Introgression and isolation contribute d to
the development of Hungarian Manga lica pigs from a particular European ancient bloodline. Genet Sel Evol **45**:22.

Martínez AM, Delgado JV, Rodero A , Vega, Pla JL
(2000) Genetic structure of
the Iberian pig breed using microsatellites. Anim Genet **31**:295-301.

Muñoz G., Óvilo C., Silió L., Tomás A., Noguera J.L. & Rodríguez M.C. (2009) Single- and joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on Sus scrofa chromosome 6 and leptin receptor effects on fatness and growth traits. Journal of Animal Science 87, 459–68.

Muñoz G., Alcázar E., Fernández A., Barragán C., Carrasco A., de Pedro E., Silió L., Sánchez J.L. & Rodríguez M.C.

(2011) Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc 9 Iberian crossbred pigs. Meat Science 88, 169–73.

Ojeda A, Huang LS, Ren J, Angiolillo A, Cho IC, Soto H, Lemús-Flores C, Makuza SM, Folch JM, Pérez-Enciso M (2008) Selection in the making: a worldwide survey of haplotypic div ersity around a causative mutation in porcine IGF2. Genetics **178**(3):1639-52.

Ojeda A, Rozas J, Folch JM, Pérez-Enciso M (2006) Unexpected high polymorphism at the FABP4 gene unveils a complex his tory for pig populations. Genetics **174**(4):2119-27.

Okumura N, Kurosawa Y, Kobayashi E, Watanobe T, Ishiguro N, Yasue H, Mitsuhashi T (2001) Genetic relationship amongst the major non-coding regions of mitochondrial DNAs in wild boars and several breeds of domesticated pigs. Anim Genet **32**(3):139-47.

Ollivier L. European pig genetic diversity: a minireview. Animal. 2009 3(7):915-24.

Ovilo C, Fernández A, Noguera JL, Barragán C, Letón R, Rodríguez C, Mercadé A, Alves E, Folch JM, Varona L, Toro M (2005) Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. Genet Res **85**(1):57-67.

Ovilo C, Oliver A, Noguera JL, Clop A, Barragán C, Varona L, Rodríguez C, Toro M, Sánchez A, Pérez-Enciso M, Silió L (2002) Test for positional candidate genes for body composition on pig chromosome 6. Genet Sel Evol **34**(4):465-79.

Ovilo C, Fernández A, Fernández AI, Folch JM, Varona L, Benítez R, Nuñez Y, Rodríguez C, Silió L (2010). Hypothalamic expression of porcine l eptin receptor (LEPR), neuropeptide Y (NPY), and cocaine- and amphetamine-regulated transcript (CART) genes is influenced by LEPR genotype. Mamm Genome **21**(11-12):583-91.

Pérez-Montarelo D, Fernández A, Folch JM, Pena RN, Ovilo C, Rodríguez C, Silió L, Fernández AI (2012) Joint effects of porcine leptin and leptin receptor polymorphisms on

productivity and quality traits. Anim Genet **43**(6):805-9.

Pérez-Montarelo D, Fernández A, Barragán C, Noguera JL, Folch JM, Rodríguez MC, Ovilo C, Silió L, Fernández AI (2013) Transcriptional Characterization of Porcine Leptin and Leptin Receptor Genes. PLoS One **8**(6):e66398.

Polzin, T. & S.V. Daneschmand (2003) On Steiner trees and minimum spanning trees in hypergraphs. Operations Research Letters **31**(1): 12-20.

Porter V (1993) Pigs. A handbook to the breeds of the world. Mountfield (United Kingdom): Helm Information Ltd.

Ren J, Duan Y, Qiao R, Yao F, Zhang Z, Yang B, Guo Y, Xiao S, Wei R, Ouyang Z, Ding N, Ai H, Huang L (2011) A missense mutation in PPARD causes a major QTL effect on ear size in pigs. PLoS Genet **7**(5):e1002043.

Rodrigáñez J, C. Barragán, E. Alves, C. Gortázar,M. A. Toro, L. Silió (2008) Genetic diversity and allelic richness in Spanish wild and domestic pig population estimated from microsatellite markers. Spanish Journal of Agricultural Research **6**: 107-115.

Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg Ö, Jern P, Jørgensen CB, Archibald AL, Fredholm M, Groenen MA, Andersson L (2012) Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci U S A **109**(48):19529-36.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J (2006) TM4 microarray software suite. Methods Enzymol **411**:134-93.

Sambrook, J., Fritsch, E.F., Maniatis, T (1989) Molecular cloning: A laboratory Manual. II ed. Cold Spring Harbour Laboratory. Press, Cold Spring Harbour, N.Y.

SanCristobal M, Chevalet C, Haley CS, Joosten R, Rattink AP, Harlizius B, Groenen MA, Amigues Y, Boscher MY, Russell G, Law A, Davoli R, Russo V,Désautés C, Alderson L, Fimland E, Bagga M, Delgado JV, Vega-Pla JL, Martinez AM, Ramos M, Glodek P, Meyer JN, Gandini GC, Matassino D, Plastow GS,Siggens KW, Laval

G, Archibald AL, Milan D, Hammond K, Cardellino R (2006) Genetic diversity within and between European pig breeds using microsatellite markers. Anim Genet **37**(3):189-98.

SanCristobal M, Chevalet C, Peleman J, Heuven H, Brugmans B, van Schriek M, Joosten R, Rattink AP, Harlizius B, Groenen MA, Amigues Y, Boscher MY,Russell G, Law A, Davoli R, Russo V, Dèsautés C, Alderson L, Fimland E, Bagga M, Delgado JV, Vega-Pla JL, Martinez AM, Ramos M, Glodek P, Meyer JN,Gandini G, Matassino D, Siggens K, Laval G, Archibald A, Milan D, Hammond K, Cardellino R, Haley C, Plastow G (2006) Genetic diversity in European pigs utilizing amplified fragment length polymorphism markers. Anim Genet **37**(3):232-8.

Sollero BP, Paiva SR, Faria DA, Guimaraes SEF, Castro STR, Egito AA, Albuquerque MSM, Piovezan U, Bertani GR, da S, Mariante A (2009) Genetic diversity of Brazilian pig breeds evidenced by microsatelite markers. Livest Sci **123**:8-15.

Steibel JP, Poletto R, Coussens PM,and Rosa GJM. (2009) A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. Genomics. 94: 146-152.

Switonski M, Stachowiak M, Cieslak J, Bartz M, Grzes M (2010) Genetics of fat tissue accumulation in pigs: a comparative approach. J Appl Genet **51**(2):153-68.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28**(10):2731-9.

Thuy NT, Melchinger-Wild E, Kuss AW, Cuong NV, Bartenschlager H, Geldermann H (2006) Comparison of Vietnamese an d European pig breeds using microsat ellites. J Anim Sci **84**(10):2601-8.

White S (2011) From Globalized pig breeds to capitalist pigs: a study in animals cultures and evolutionary history. Environmental History **16**: 94-120.

Wilkinson S, Lu ZH, Megens HJ, Archibald AL, Haley C, Jackson IJ, Groenen MA, Crooijmans RP, Ogden R, Wiener P (2013) Signatures of

diversifying selection in European pig breeds. PLoS Genet **9**(4):e1003453.

Wu GS, Yao YG, Qu KX, Ding ZL, Li H, Palanichamy MG, Duan ZY, Li N, Chen YS, Zhang YP (2012) Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multipledome stication events in East Asia. Genome Biol. 2007;8(11):R245.

Yang M, Yang B, Yan X, Ouyang J, Zeng W, Ai H, Ren J, Huang L. Nucleotide variability and linkage dis equilibrium patterns in the porcine MUC4 gene. BMC Genet **13**:57.

Wylie ARG (2011) Leptin in farm animals: where are we and where can we go? Animal **5**:246-267.

TABLES

Table 1. Animal material used in the present study, number of pigs (N) number (S) and proportion (Ps) of segregating sites, nucleotide diversity index (π), number of haplotypes (h) and private haplotypes (ph) and allelic richness (hr) of each group.

Population			Nucleotide variation		Haplotypic variation		
	2N	S	P_S	π	H	Hr	ph
LEP gene							
Spanish Wild Boar	18	37	0.552	0.197 (±)	17	9.7	13
Iberian Breed	24	36	0.529	0.197 (±)	18	8.8	14
Other European Breeds	40	41	0.612	0.237(±)	30	9.7	26
Asian Domestic Breeds	10	55	0.821	0.380 (±)	9	9	9
TOTAL	92	68			68	37.2	
LEPR gene							
Spanish Wild Boar	18	18	0.0150	0.0044(±0.0009)	7	5.6	5
Iberian Breed	24	0	0	0	1	1	0
Other European Breeds	40	20	0.0166	0.0050(±0.0010)	10	4.7	6
Asian Domestic Breeds	10	10	0.0083	0.0038(±0.0012)	3	3	1
TOTAL	92	24	0.0039	0.004(±0.0010)	16	14.3	

Table 2. Mean genetic distances between the clusters detected in the *LEP* (**a**) and *LEPR* (**b**) genes analyses. The values in the diagonal, highlighted in grey color, show the average within-groups genetic distance, the values below the diagonal the net between groups genetic distance and the values above the diagonal the between groups genetic distances.

a)

	European (A)	**European (B)**	**Vietnamese (C)**	**Chinese (D)**
European (A)	0.13(±0.02)	0.36(±0.05)	0.40(±0.05)	0.48(±0.04)
European (B)	0.25(±0.04)	0.09(±0.01)	0.39(±0.05)	0.47(±0.04)
Vietnamese (C)	0.29(±0.05)	0.30(±0.05)	0.09(±0.02)	0.54(±0.04)
Chinese (D)	0.29(±0.04)	0.30(±0.05)	0.38(±0.05)	0.24(±0.03)

b)

	European	**Asiatic**
European	0.22(±0.04)	0.49(±0.06)
Asiatic	0.26(±0.07)	0.22(±0.05)

Figure 1. Neighbour-joining dendrogram of the 28 animals included in the *LEP* analyses using the p-distance computed using the detected haplotypes.

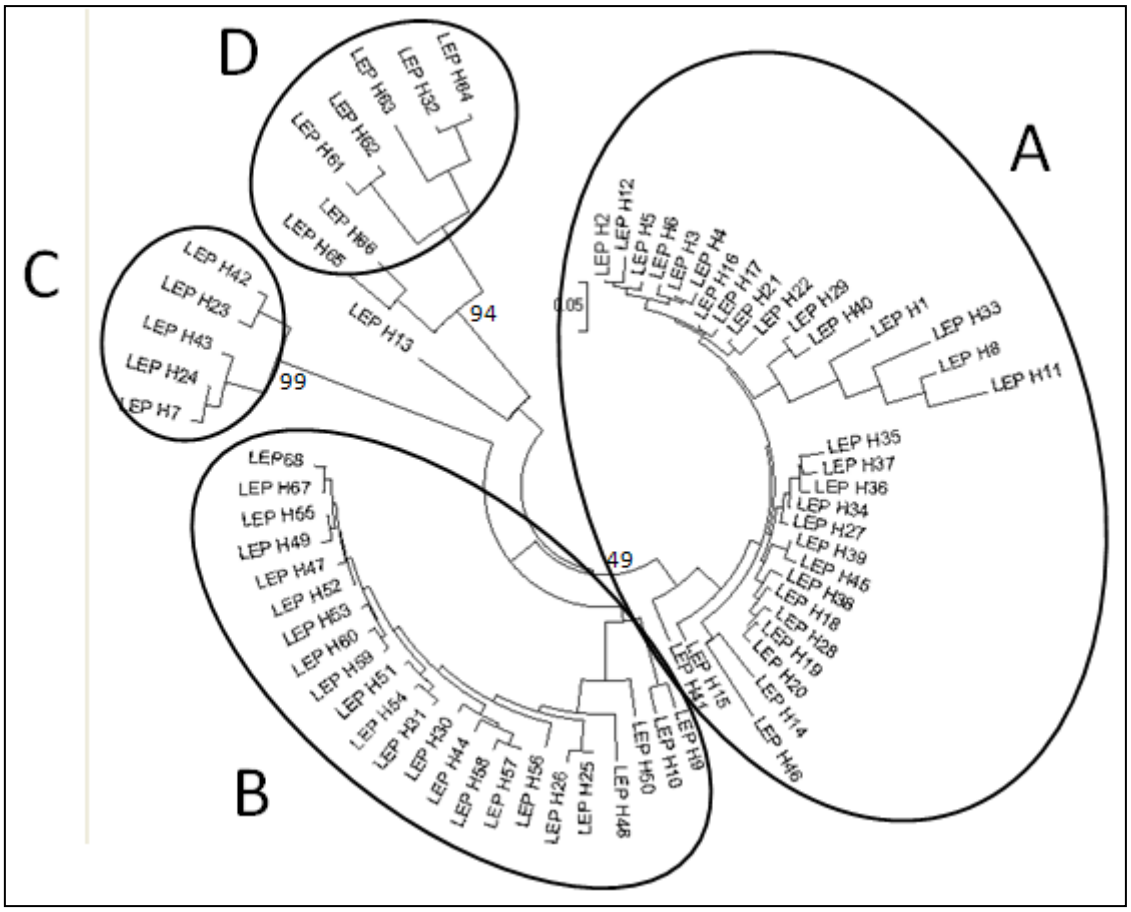


Figure 2. Heatmap of the pair-wise genetic distances between *LEP* haplotypes. Shade scale is shown below the heatmap; darker color indicates smaller distances.

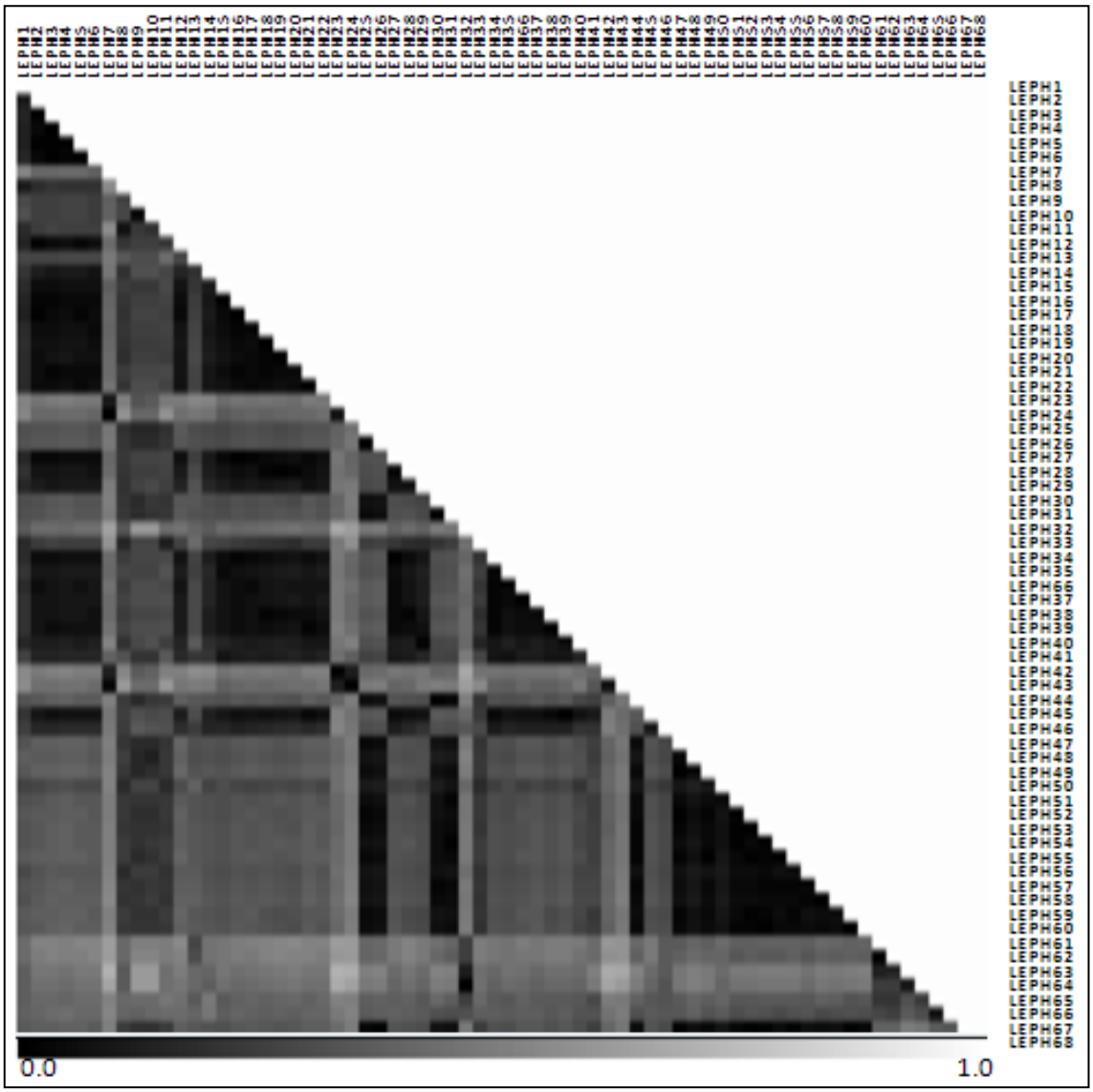


Figure 3. Median-joining network obtained with the 40 haplotypes of *LEP* gene. The area of the circles is proportional to the haplotypes frequency. In the network, Asiatic haplotypes are represented in pink, European wild boars haplotypes in green, European domestic pig haplotypes in blue and Iberian haplotypes in orange.

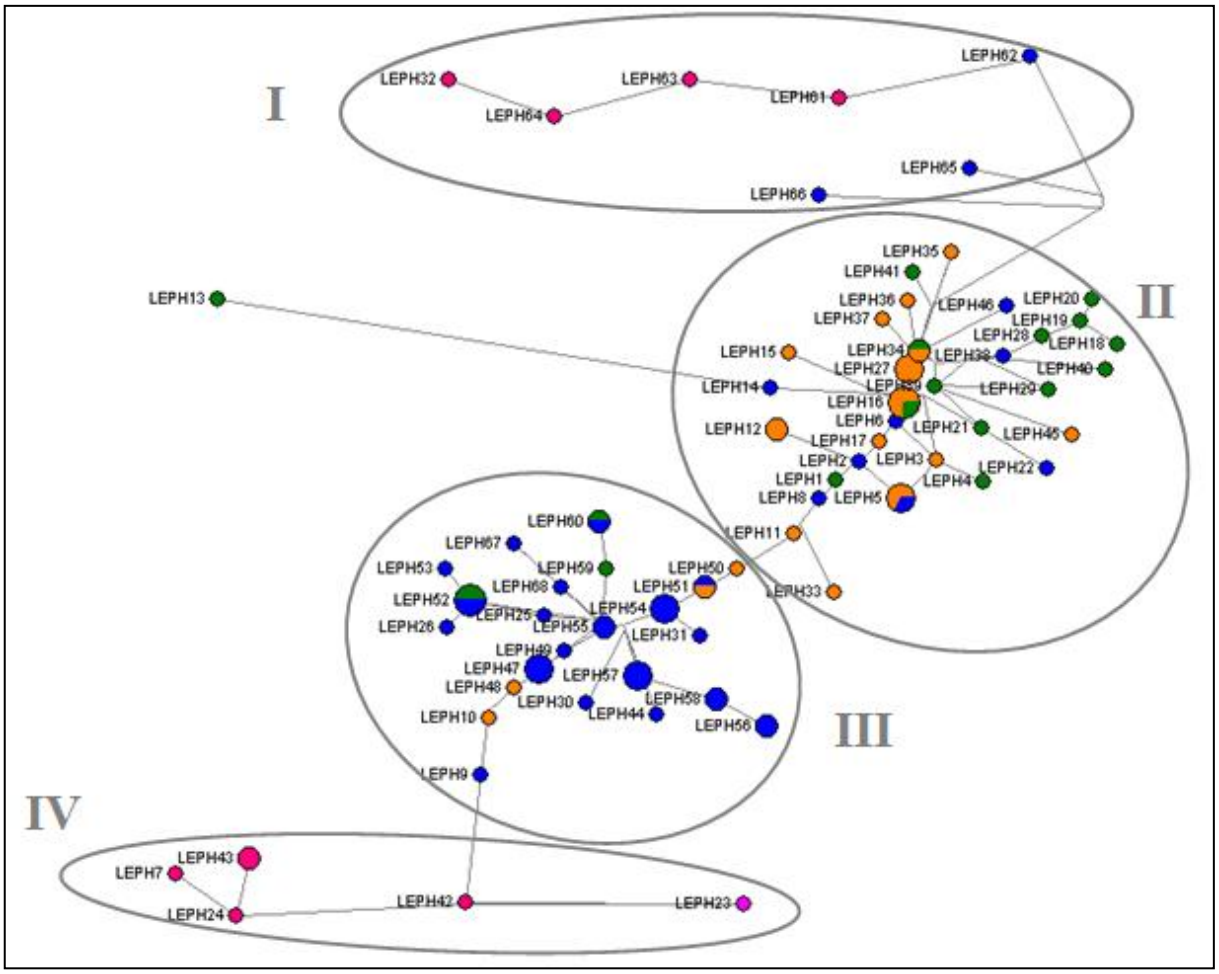


Figure 4. Neighbour-joining dendrogram using the p-distance computed using the *LEPR* gene haplotypes of the 46 analyzed animals.

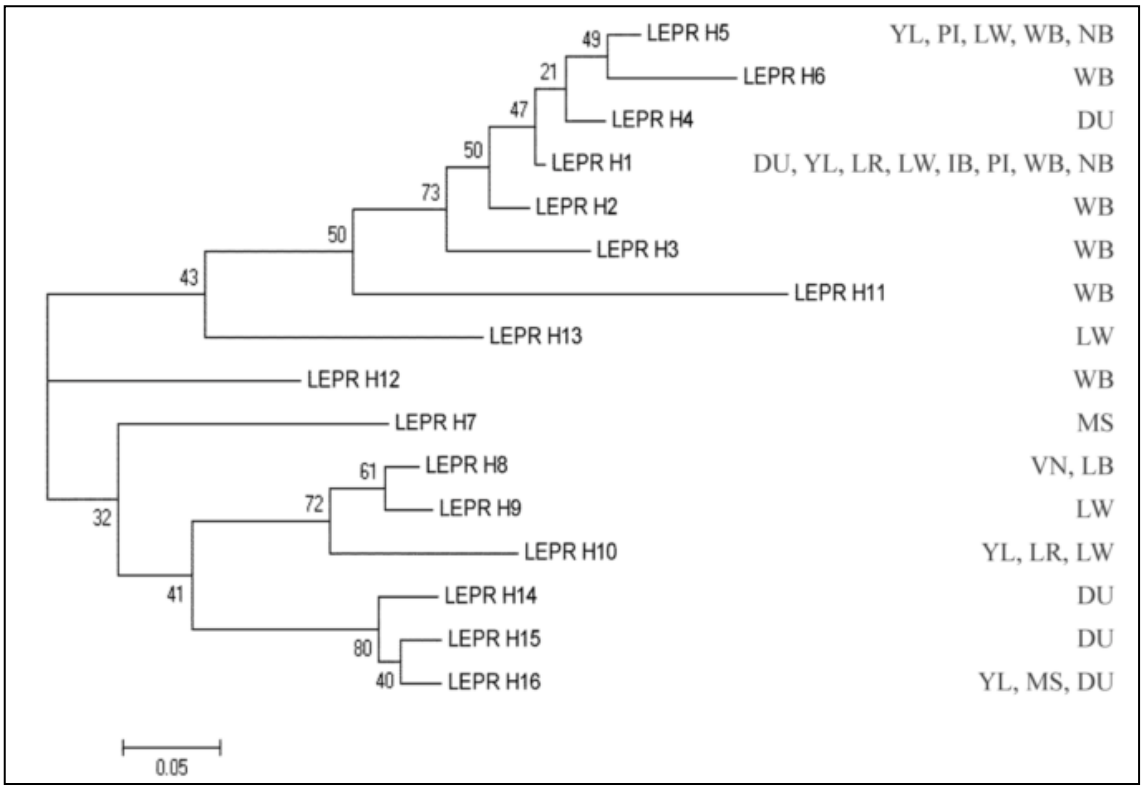


Figure 5. Heatmap of the pair-wise genetic distances between *LEPR* haplotypes. Shade scale is shown below the heatmap; darker color indicates smaller distances.

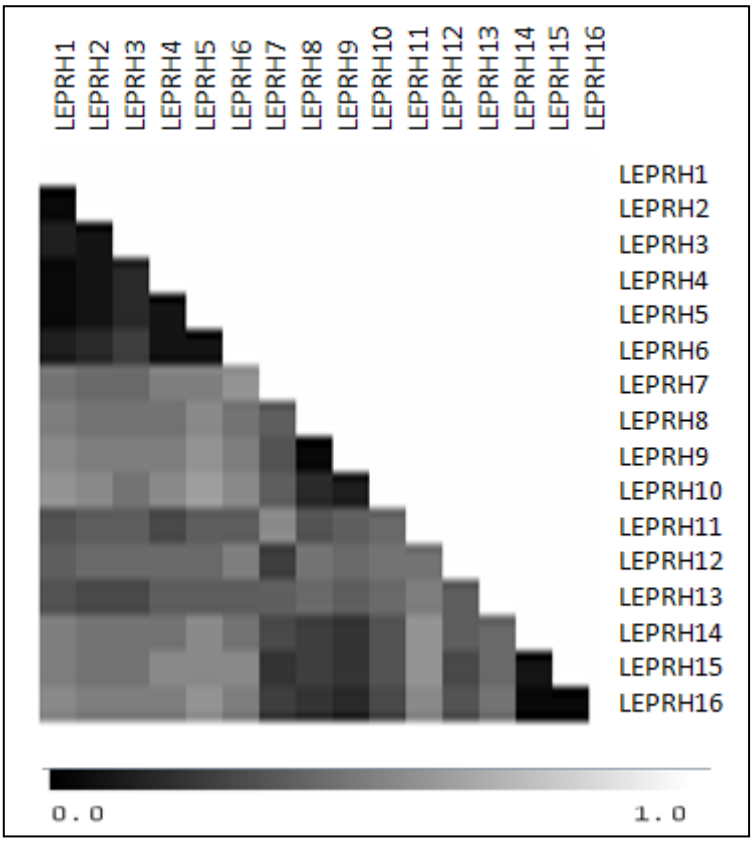
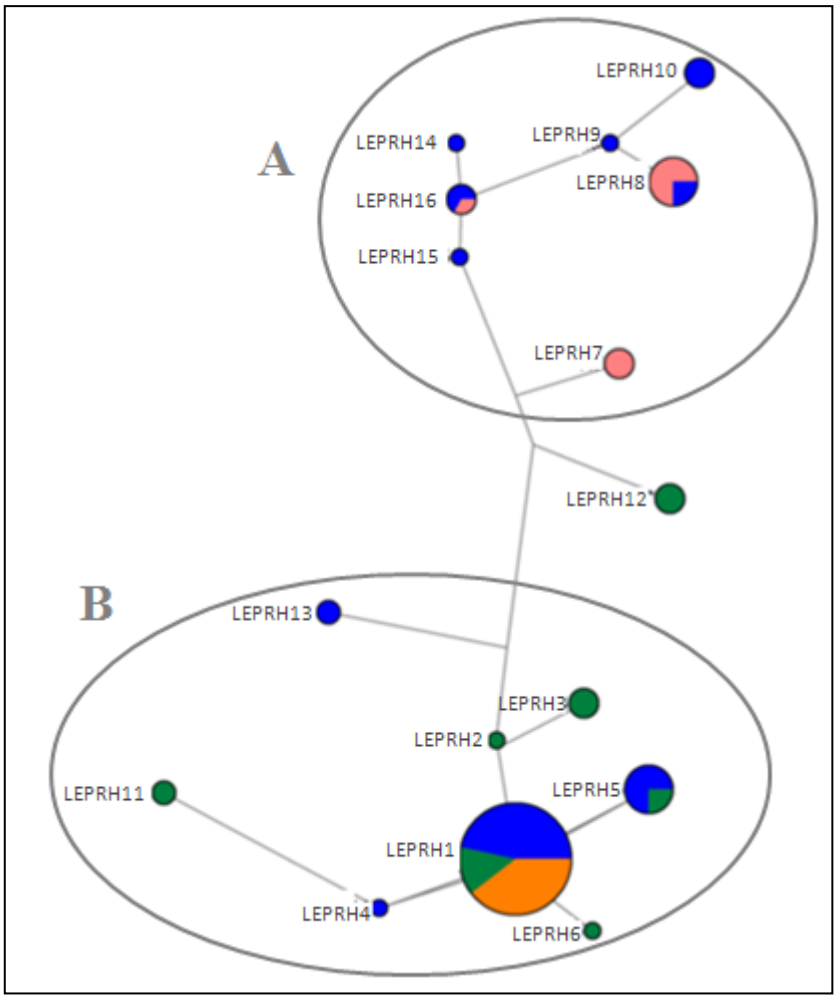


Figure 6. Median-joining network obtained with the 16 haplotypes of *LEPR* gene. The area of the circles is proportional to the haplotypes frequency. In the network, Asiatic haplotypes are represented in pink, European wild boars haplotypes in green, European domestic pig haplotypes in blue and Iberian haplotypes in orange.



ARTÍCULO V

Transcriptional characterization of porcine *leptin* and *leptin receptor* genes

OPEN ACCESS Freely available online

PLOS ONE

Transcriptional Characterization of Porcine *Leptin* and *Leptin Receptor* Genes

Dafne Pérez-Montarelo[1]*, Almudena Fernández[1], Carmen Barragán[1], Jose L. Noguera[2], Josep M. Folch[3,4], M. Carmen Rodríguez[1], Cristina Óvilo[1], Luis Silió[1], Ana I. Fernández[1]

1 Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain, **2** Genètica i Millora Animal, Institut de Recerca i Tecnologies Agroalimentaries, Lleida, Spain, **3** Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Spain, **4** Genètica Animal, Centre de Recerca en Agrigenòmica, Bellaterra, Spain

Abstract

The leptin (LEP) and its receptor (LEPR) regulate food intake and energy balance through hypothalamic signaling. However, the LEP-LEPR axis seems to be more complex and its expression regulation has not been well described. In pigs, LEP and LEPR genes have been widely studied due to their relevance. Previous studies reported significant effects of SNPs located in both genes on growth and fatness traits. The aim of this study was to determine the expression profiles of LEP and LEPR across hypothalamic, adipose, hepatic and muscle tissues in Iberian x Landrace backcrossed pigs and to analyze the effects of gene variants on transcript abundance. To our knowledge, non porcine LEPR isoforms have been described rather than LEPRb. A short porcine LEPR isoform (LEPRa), that encodes a protein lacking the intracellular residues responsible of signal transduction, has been identified for the first time. The LEPRb isoform was only quantifiable in hypothalamus while LEPRa appeared widely expressed across tissues, but at higher levels in liver, suggesting that both isoforms would develop different roles. The unique LEP transcript showed expression in backfat and muscle. The effects of gene variants on transcript expression revealed interesting results. The LEPRc.1987C>T polymorphism showed opposite effects on LEPRb and LEPRa hypothalamic expression. In addition, one out of the 16 polymorphisms identified in the LEPR promoter region revealed high differential expression in hepatic LEPRa. These results suggest a LEPR isoform-specific regulation at tissue level. Conversely, non-differential expression of LEP conditional on the analyzed polymorphisms could be detected, indicating that its regulation is likely affected by other mechanisms rather than gene sequence variants. The present study has allowed a transcriptional characterization of LEP and LEPR isoforms on a range of tissues. Their expression patterns seem to indicate that both molecules develop peripheral roles apart from their known hypothalamic signal transduction function.

Citation: Pérez-Montarelo D, Fernández A, Barragán C, Noguera JL, Folch JM, et al. (2013) Transcriptional Characterization of Porcine *Leptin* and *Leptin Receptor* Genes. PLoS ONE 8(6): e66398. doi:10.1371/journal.pone.0066398

Editor: Raul M. Luque, University of Cordoba, Spain

Received March 6, 2013; **Accepted** May 6, 2013; **Published** June 18, 2013

Funding: This work was funded by Ministerio de Ciencia e Innovación (MICINN) projects AGL2008-04818-C03 and AGL2011-29821-C02-02. D. Pérez-Montarelo was funded by a Formación de Personal Investigador (FPI) PhD grant from the Spanish Ministerio de Ciencia e Innovación (BES-2009-025417). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dafne.perez@inia.es

Introduction

The leptin hormone, coded by the *LEP* gene, regulates energy balance, food intake and body weight [1]. Leptin is mainly secreted by white adipocytes into the blood stream. At hypothalamic level, it interacts with its receptor *LEPR* that encodes a signal transductor that activates the Janus kinases (JAK) and signal transducers and activators of transcription (STAT) as a response to leptin. Overall, these processes result in an increase of energy expenditure and physical activity and a reduction of the food intake, driving an adipose mass reduction [2]. However the role of the LEP-LEPR axis has been shown to be more complex. There are evidences revealing that leptin may act peripherally as well [3]. Although a single *LEP* isoform has been described in most mammals, six different *LEPR* isoforms have been identified in several species. Both, in human and mouse, the longer *LEPRb* isoform is mainly expressed in hypothalamus [4,5] and appears to be the dominant signaling isoform regulating food intake and body weight [6,7,8,9]. In human, expression of the short *LEPRa* isoform has been detected in most tissues; supposedly involved in leptin transport through the blood-brain barrier or in leptin degradation [10]. Indeed, it has been shown that in many tissues short *LEPR* isoforms predominate [11]. Nevertheless, the specific function and tissue distribution of short *LEPR* isoforms remain unclear in all species.

In pigs, *LEP* and *LEPR* genes have been widely studied due to their relevance on important economic traits such as growth and fatness [12,13]. Our previous studies on an Iberian x Landrace experimental cross reported significant effects of SNPs located in both genes on pig productive traits [14,15,16]. A highly significant and strong additive effect on fatness and growth has been reported for *LEPR*c.1987C>T polymorphism in this population [14]. Moreover, the effect of this SNP on growth and fatness has been confirmed in very different genetic backgrounds (crossbred Iberian x Meishan, Duroc x Iberian and Duroc x Landrace/Large White pigs) and growth stages [17–20]. In addition, differential *LEPRb* expression according to this SNP was found in hypothalamus [15]. In the same Iberian x Landrace experimental intercross, effects of *LEP*g.1387C>T and joint effects of *LEP*g.1387C>T and *LEPR*c.1987C>T polymorphisms were detected on growth,

fatness, body composition and fatty acid composition [16]. Despite their relevant role, the expression patterns of porcine *LEP* and *LEPR* isoforms across tissues as well as gene expression regulation have not been characterized so far.

The objective of the present study was to improve the characterization of porcine *LEP* and *LEPR* genes in order to better understand their potential biological roles. A search of short *LEPR* isoforms has been carried out and expression of *LEP* and *LEPR* isoforms has been characterized across five different tissues (backfat, liver, hypothalamus, *Longissimus dorsi* and diaphragm) in a backcross (BC) of Iberian x Landrace pigs. Previously, a set of reference genes has been tested in order to establish the most adequate control genes to evaluate *LEP* and *LEPR* gene expression differences across the five tissues. In addition, the variation of *LEP* and *LEPR* promoter regions and the differential expression of both genes conditional on several SNPs genotypes have also been investigated.

Materials and Methods

Ethics Statement

Animal manipulations were performed according to the Spanish Policy for Animal Protection RD1201/05, which meets the European Union Directive 86/609 about the protection of animals used in experimentation. Research protocols were approved by Animal Care and Use Committee of the Institut de Recerca i Tecnologies Agroalimentaries. The animals used in the present study grown in an experimental farm in good conditions and were fed according to their necessities. Electric stunning was used to ameliorate the suffering of the animals before sacrifice. The animals sacrifice took place at the PRIMAYOR slaughter-house, property of the company "Primayor Foods S.L." in Lleida, Spain. All the animals used and their information belong uniquely to the project AGL2011-29821-C02-02, and the slaughterhouse has no responsibility for them.

Animal Material, DNA and RNA Extraction and cDNA Synthesis

Animals used in this study belong to a BC generated from the IBMAP population [21]. In brief, three Iberian boars were mated to 30 Landrace sows (F0) to produce 70 F1 animals. The BC was generated by mating 5 F1 boars with 25 Landrace sows to produce 187 BC animals. A total of 40 BC males belonging to the same batch were selected to identify new isoforms and to measure *LEP* and *LEPR* genes expression. Genomic DNA from parental pigs and the 40 selected BC animals was extracted from blood samples with a standard phenol: chloroform protocol, and used for promoter sequencing and polymorphisms genotyping. Samples of liver, hypothalamus, *Longissimus dorsi*, diaphragm and backfat (taken at the level of the fourth rib) from the BC animals were collected at slaughter at an average age of 179.9±2.6 days and 98.46±14.32 kg, immediately frozen in liquid nitrogen and stored at −80°C until analyzed. Total RNA was extracted from the five tissues using the RiboPure kit (Ambion, Austin, TX, USA) following the manufacturer's recommendations. The total RNA was quantified using NanoDrop-100 spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA). The integrity of the RNA was assessed using an Agilent 2100 Bioanalyzer device (Agilent technologies, Santa Clara, CA, USA). The cDNA synthesis was performed using the Superscript II enzyme (Invitrogen, Carlsbad, CA, USA) with random hexamers in a total volume of 20 µl containing 1.5 µg of total RNA, following the supplier's instructions.

Identification of Novel Porcine *LEPR* Isoforms

A total of six different *LEPR* isoforms have been detected in mouse and human. However, only the longest one (*LEPRb*) has been described in pig so far [1]. The identification of short *LEPR* isoforms conducted in the present study was based on the homology between human *LEPR* isoforms (GenBank: NG_015831.2) and porcine (GenBank: FN677933.1) *LEPR* gene sequence. Primer pairs were designed on highly conserved regions between human and pig (Table S1) for those isoforms containing alternative exons. The primer pair named LEPR19-20′ was designed to amplify a putative porcine short *LEPR* isoform, homologous to human short isoform *LEPR*a (also known as LEPR-002 at Ensembl). This short isoform is identical to the *LEPRb* except for a shorter alternative exon 20 (named exon 20′) (Figure 1). Another primer pair, called LEPR5-20 was designed to amplify a shorter isoform homologous to the human LEPR-202 isoform (Ensembl). In human, *LEPR*-202 isoform is identical to *LEPRb* from exons 1 to 5 and 20 but lacking the segment covering exons 6 to 19 (Figure 1). PCRs were carried out, in four samples of each analyzed tissue, in a 25 µl final volume containing 2.5 µl of cDNA, 1 unit of Taq polymerase (Biotools), specific buffer, 2.5 mM of dNTPs and 0.5 µM of each primer. The specific annealing temperature of each primer pair is shown in Table S1.

The ExPASy tool [22] was used to translate the mRNA sequence to protein and SMART tool [23] to predict differences in the protein domains between the isoforms.

Promoters Sequence Analyses and Polymorphisms Identification

Genomic DNA from the parental animals of the IBMAP population, three Iberian boars and 30 Landrace sows, was used for sequencing the *LEP* and *LEPR* promoter regions. The Promoter Scan tool, PROSCAN version 1.7 [24], and Promoter Inspector software [25], were used to validate putative promoter regions in the 5′UTR of *LEP* and *LEPR* genes, in order to confirm the previously described location of the promoters of both genes. Two primer pairs, LEPpro1 and LEPpro2 (Table S1), were designed to amplify 900 bp, in two overlapped fragments, of the 5′ *LEP* region in the promoter described by Stachowiak et al. [26] according to the available sequence GenBank AF492499.2. *LEPR* promoter region sequencing was conducted in accordance with the pig *LEPR* promoter described by Lee et al [27]. According to the available *LEPR* gene sequence (GenBank: FN677933.1), three primer pairs (Table S1) were designed to amplify 1,266 bp, in three overlapped fragments, of the 5′ region of this gene. PCRs were carried out in a 25 µl final volume containing 100 ng of DNA, 1 unit of polymerase (Biotools) or HotStart polymerase (Quiagen), specific buffer, 2 mM of dNTPs and 0.5 µM of each primer. The specific annealing temperature of each primer pair is shown in Table S1. The PCR reactions were carried out in a GeneAmp PCR System 9700 (Applied Biosystems, Warrington, UK). The PCR products from the *LEP* gene were purified using the ExoSAP-IT® method (Affymetrix) and the *LEPR* promoter PCR products with the GFX™ PCR DNA purification kit (GE Healthcare, UK) according to the manufacturers' protocol. All products were sequenced with both forward and reverse primers using the 3100 BigDye® Terminator v3.1 Matrix Standard in a 3730 DNA Analyzer (Applied Biosystems Warrington, UK). The obtained sequences were edited and aligned using the EditSeq and MegAlign packages of the WinStar software for the identification of polymorphisms.

Transcription factors (TF) binding sites were examined in both promoter regions using the Molecular Informatics Resource for the Analysis of Gene Expression website of the Institute for

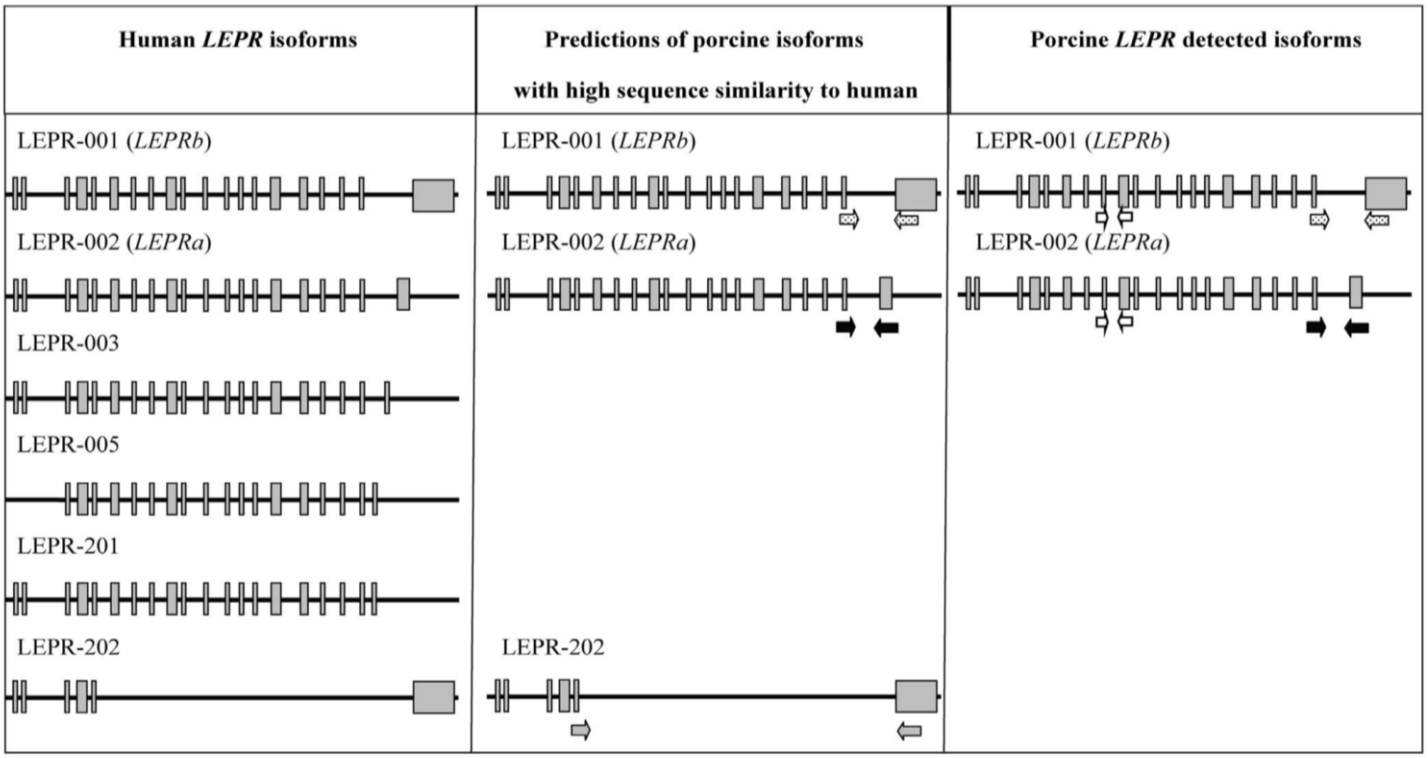


Figure 1. Schematic representation of the *LEPR* isoforms. Human *LEPR* isoforms, predictions of porcine *LEPR* isoforms showing high sequence similarity to human and detected porcine isoforms. The arrows represent the primers location for isoforms identification on the predicted porcine isoforms and for isoforms quantification on the porcine detected isoforms. Empty arrows represent *LEPRglobal* primers, arrows filled with black dots represent *LEPRb* primers, black filled arrows *LEPRa* primers and grey filled arrows primers designed to detect LEPR-202 isoform.
doi:10.1371/journal.pone.0066398.g001

Transcriptional Informatics (IFTI Mirage website) [28]. The search was made with the IFTI Tfsite option and mammalian sites as query parameters.

Gene Expression Quantification

Relative transcript quantification of samples from liver, *Longissimus dorsi*, diaphragm, hypothalamus and backfat samples of the 40 selected males was performed in 384 plates using the LightCycler®480 Real-Time PCR System (Roche Diagnostic, Mannheim, Germany). Real-time qPCR reactions were performed in a total volume of 20 µl containing 2.5 µl of cDNA (1/10 dilution), 10 µl of Roche LightCycler mix and a specific amount of primer pairs (0.3 µl for reference genes, 0.4 µl for *LEPRglobal* and 0.6 µl for *LEPRb*, *LEPRa* and *LEP* genes, at 5 µM dilution in all cases). All primer pairs used are detailed in Table S1. Standard PCR on cDNA were carried out to verify amplicon sizes. A non-template control, without cDNA, was included as negative control. Cycling conditions were 95°C for 10 min, followed by 45 cycles of 95°C (15 sec) and 60°C (1 min), when the fluorescence was acquired. Finally, a dissociation curve to test PCR specificity was generated by one cycle at 95°C (15 sec), followed by 60°C (20 sec) and ramped up to 95°C with the fluorescence acquired during the increase to 0.01°C/sec. Data were analyzed with LightCycler 480 software (Roche) using the second derivative method [29]. All points and samples were run in triplicates as technical replicates and dissociation curves were analyzed for each individual replicate. Single peaks in the dissociation curves confirmed the specific amplification of the primer pairs and the absence of primer dimers. PCR efficiency was estimated by standard curve calculation using four points of cDNA serial dilutions (1:2, 1:4 and 1:8) of a pool of five samples (one from each tissue).

Reference Genes Selection

A set of six genes *(GADPH, TBP, TOP2B, B2M, ACTB* and *eEF2*) commonly used as reference genes in porcine expression studies were selected from the literature [30,31]. These genes are involved in different biological processes and functions and were selected to avoid genes belonging to the same pathways that may be co-regulated, because the geNorm algorithm used assumes no co-regulation of housekeeping genes. Their stability was evaluated across the five analyzed tissues. The primer pairs used were described in Kuijk et al. [32] and Erkens et al. [33], except for *eEF2* gene for which primers were designed according to AK240374 sequence (Table S1). A total of ten samples, two from each tissue, were used to measure the stability of the reference genes across tissues. The gene stability measures (M) were calculated using geNorm algorithm [34]. Those genes with the lowest M values have the most stable expression in each particular condition.

LEP and *LEPR* Gene Expression Quantification

The relative expression measures of *LEP* and *LEPR* transcripts were determined in the five tissues for the 40 selected backcrossed pigs. Primers for *LEP* mRNA quantification were designed according to the available sequence (GenBank NM_213840.1) covering exons 2 and 3 (Table S1). Two different isoforms of the *LEPR* gene were analyzed: the *LEPRb* and the shorter one described here for the first time (*LEPRa*). Primer pairs were designed from the GenBank AF092422.1 porcine sequence between exons 18 and 20 for the *LEPRb*, and between exons 19 and 20′ for the *LEPRa* (Table S1). Moreover, the total *LEPR* expression (*LEPRglobal*) was measured by a primer pair designed

between exons 8 and 9 which are shared by all isoforms detected so far (Table S1).

Genotyping of Polymorphisms

The previously reported *LEPR*c.1987C>T and *LEP*g.1387C>T polymorphisms and the polymorphisms detected in the present study in *LEP* and *LEPR* promoter regions were genotyped in the 40 BC animals selected for the gene expression analyses. The *LEPR*c.1987C>T and *LEP*g.1387C>T polymorphisms genotypes were determined on DNA samples by pyrosequencing technology following the protocols described in Óvilo et al. [14] and Pérez-Montarelo et al. [16], respectively. The genotyping of the polymorphisms detected in *LEP* and *LEPR* promoter regions was conducted by sequencing, using the same primer pairs and conditions used for polymorphisms identification. Additionally, 11 females from the same BC were genotyped and included in the differential *LEPR* gene expression analysis conditional on *LEPR*g.35856G>A genotype. The Haploview software was used to estimate the linkage disequilibrium between the identified SNPs per gene [35].

Differential Expression Analyses

Statistical analysis of gene expression was carried out following the method proposed by Steibel et al [36], which consists of the analysis of cycles to threshold values (C_p), for the targets and housekeeping genes using a linear mixed model. The following model was used for analyzing the joint expression of the target (*LEP* and *LEPR*) and control (*ACTB* and *B2M*) genes in different tissues:

$$y_{gijkm} = TG_{ig} + P_j + S_k + B_{gim} + D_{im} + e_{gijkm}$$

where $y = \log_2(E_{gijkm} - Cpgijkm)$, E is the efficiency of the PCR of each gene, *Cp* is the mean value obtained from the thermocycler software from the three replicates of *g*th gene in the *k*th plate in a sample collected from the *i*th tissue of the *m*th animal, P_j and S_k are the systematic effects of *j*th plate and *k*th gender (only for the differential *LEPR* expression analysis conditional on *LEPR*g.35856G>A genotype, where 11 females were included), TG_{ig} is the specific effect of tissue *i* on the expression of gene *g*, B_{gim} is a gene-specific random effect of the *m*th pig on the *i*th tissue, D_{im} is a random tissue sample-specific effect common to all the genes, and e_{gijkm} is a residual effect. A similar model was used for estimating the expression rates in each tissue of different genotypes of *LEP* and *LEPR* genes replacing TG_{ig} by TG_{hig}, which is the specific effect of genotype *h* on the expression of gene g in tissue *i*. Homogenous residual variances were assumed in these models according to the results of preliminary analyses with models fitting heterogeneous variances. Heteroscedasticity was discarded because of the small estimated differences among residual gene-specific variances and their marginal effects on the differential expression tests.

To test differences in the expression rate of genes of interest ($diff_{TG}$) between classes (alternative tissues or genotypes) normalized by the housekeeping genes (*HK*), different contrasts were performed between the respective estimates of *TG* levels. Significance of $diff_{TG}$ estimates was determined with the *t* statistic. To obtain fold change values from de estimated $diff_{TG}$ values, the following equation was applied: $FC = 2^{-diff_{TG}}$. Asymmetric 95% confidence intervals (CI) were calculated for each *FC* value by using the standard error (SE) of the estimated difference: 95% CI from $2^{-(diffTG+1.96xSE)}$ to $2^{-(diffTG-1.96xSE)}$.

Results

Identification of a Novel Porcine *LEPR* Isoform (*LEPRa*)

In order to identify novel porcine short *LEPR* isoforms two different primer pairs were designed: *LEPR*19-20′ and *LEPR*5-20. The shorter isoform homologous to human *LEPR*-202 could not be amplified (*LEPR*5-20 primer pair). However, a novel porcine *LEPR* isoform (*LEPRa*), homologous to human and murine *LEPRa*, was identified with LEPR19-20′ primer pair (Figure 1). This isoform is identical to the *LEPRb* described in pig from exons 1 to 19, but contains a different exon 20, probably shorter than the exon 20 of the *LEPRb*, according to its human homologous *LEPRa*. The predicted protein coded by the *LEPRb* contains 1,165 aminoacids while the protein coded by the *LEPRa* contains only 897. Even though the *LEPRb* is longer, the SMART predictions indicated that there were no differences in the domains present in both isoforms. According to this tool, both *LEPRb* and *LEPRa* have three FN3 (fibronectin type 3) domains and one transmembrane domain, all located in the region shared by both isoforms.

LEP and *LEPR* Promoter Analyses

A total of 1,266 bp of the *LEPR* promoter region (positions 34669–35935 in the reference sequence GenBank: FN677933.1) were sequenced in the 40 BC animals used in the present study. This fragment contains the *LEPR* putative promoter, non-coding exons 1 and 2 and surrounding regions. The Promoter Scan and Promoter Inspector software confirmed the location of the predicted *LEPR* promoter within the analyzed region. A total of 16 polymorphisms were identified in this fragment (Table S2), 13 previously identified by Lee et al, [27] and three new ones (*LEPR*g.35782.indelGGAGGCCCCCGGGGCGA, *LEPR*-g.35805A>G and *LEPR*g.35856). A total of 14 out of the 16 polymorphisms detected on the *LEPR* promoter were located within predicted TF binding sites ($p<0.10$) according to the IFTI Mirage website (Table S2).

A total of 900 bp of the *LEP* promoter region were also sequenced in the 40 animals. This fragment contains the *LEP* promoter, non-coding exon 1 and surrounding regions. The *LEP* promoter predicted by Promoter Scan agreed with the one reported by Stachowiak et al. [26] located within the analyzed region. A total of seven polymorphisms were identified in this region (Table S2). The search of potential transcription factors binding sites conducted with the IFTI Mirage website revealed that five of the seven detected polymorphisms were located within predicted TF binding sites ($p<0.10$) (Table S2).

Gene Expression Quantification

Selection of the most suitable reference genes. In order to find the most stable reference genes to normalize gene expression measures across the five porcine tissues, the stability of six commonly used reference genes was tested: *GADPH, TBP, TOP2B, B2M, ACTB* and *eEF2*. The *eEF2* gene was discarded for further analyses due to amplifications problems. The stability (M values) provided by the geNorm software for the remaining five genes are represented in Figure S1. Three of the tested genes showed M values below 1.1 (*B2M, ACTB* and *TOP2B*). In order to choose the best pair of genes among those three most stable genes, their PCR efficiencies were taken into account (93%, 80% and 87% for *B2M, ACTB* and *TOP2B*, respectively). According to the M value and PCR efficiency, *B2M* and *ACTB*, were selected as the most suitable reference genes in our experiment, and were used for expression data normalization.

Porcine *LEP* and *LEPR* Genes Expression Across Tissues

Porcine *LEP* and *LEPR* genes expression levels showed a wide variation across the tested tissues (hypothalamus, backfat, liver, *Longissimus dorsi* and diaphragm). The results obtained for the global *LEPR* expression (*LEPRglobal*), that includes both detected isoforms (*LEPRa* and *LEPRb*), as well as, other isoforms that could potentially exist, are shown in Figure 2A (relative to the tissue that showed the lowest expression, backfat in this case) (Table S3). The highest *LEPRglobal* expression was found in liver followed by hypothalamus, muscles and finally backfat. Significant expression differences were detected among all tissues in the pairwise comparisons, except between both muscles (p-value = 0.37).

In addition, specific *LEPR* isoforms *(LEPRb* and *LEPRa)* expression differences were also measured. The *LEPRb* isoform showed very low expression levels, even undetectable, in backfat, liver, *Longissimus dorsi* and diaphragm and therefore could not be quantified in these tissues. This isoform could only be quantified in hypothalamus [15]. Conversely, *LEPRa* isoform expression could be detected in all the analyzed tissues. The significant differences of the *LEPRa* isoform expression across tissues are shown in Figure 2B in pairwise comparisons, relative to backfat, that showed the lowest *LEPRa* expression (Table S3). The highest *LEPRa* expression was detected in liver, followed by muscles, hypothalamus and backfat, which showed 15 times lower expression than liver. Significant differences were detected between liver and the remaining tissues ($p<0.0001$) and between backfat and the rest of the tissues ($p<0.005$).

The *LEP* expression analyses revealed large differences among tissues. This gene showed very low, almost undetectable, expression in hypothalamus and liver, therefore it could not be quantified in those tissues. Figure 2C represents the *LEP* unique isoform expression measures on backfat and *Longissimus dorsi* relative to diaphragm that showed the lowest expression (Table S3). LEP expression in backfat was more than 100 and 56 times higher than in diaphragm ($p<0.0001$) and *Longissimus dorsi*, respectively. Smaller, but still significant, were the differences detected between both muscles ($p<0.05$).

Differential Expression Conditional on Genotypes

The expression differences of the different transcripts (*LEPRglobal*, *LEPRb*, *LEPRa*, and *LEP*) conditional on the genotypes of the previously analyzed polymorphisms, *LEPR*c.1987C>T and *LEP*g.1387C>T [14,16], and the ones identified here in the promoter regions of both genes, were tested in the tissues where they showed quantifiable expression (Tables 1 and 2).

LEPR expression conditional on genotypes of LEPRc.1987C>T and LEPR promoter SNPs. The Haploview software showed that the *LEPR* promoter SNPs are highly linked (mean $r^2 = 0.6$) (Figure 3A) and four groups of cosegregating SNPs could be identified (Table S2). Conversely, the *LEPR*c.1987C>T, located more than 65Kb far away from the promoter, and the promoter SNPs showed very low linkage (mean $r^2 = 0.06$) (Figure 3A). Significant *LEPRglobal* differential expression was found in backfat ($p<0.05$) and liver ($p<0.005$) conditional on *LEPR*c.1987C>T genotype, where the C allele in both tissues increases the *LEPRglobal* gene expression (Table 1). The specific *LEPRb* expression conditional on *LEPR*c.1987C>T genotype in hypothalamus, was previously investigated in the same animal material, showing a higher *LEPRb* expression associated also to the C allele [15]. Similar results has been found in the present study (Table 1) using a more complex statistical model. Regarding *LEPRa* isoform, also significant differential expression was found in backfat ($p<0.01$) and liver ($p<0.05$) according to *LEPR*c.1987C>T genotype (Table 1). In both tissues, the C allele showed a higher expression than T, as shown in previous *LEPRglobal*. An effect of this SNP on *LEPRa* has also been detected in hypothalamus; however, the effects seem to differ from the one reported for the *LEPRb* isoform in hypothalamus. In hypothalamus, whereas the C allele is associated with a higher *LEPRb* expression, it results in a lower expression of *LEPRa*, suggesting a possible isoform specific regulation in this tissue. No significant differential expression was detected in muscle tissues for any isoform according to this SNP.

To test whether the polymorphisms identified in the promoter region of *LEPR* gene play a role in gene expression regulation, expression differences of *LEPRglobal*, *LEPRb* and *LEPRa*, conditional on these genotypes were also tested. The 16 polymorphisms identified in the *LEPR* promoter region were classified in four groups of cosegregating SNPs (Table S2) in the 40 backcrossed animals. The SNPs *LEPRg*.34996C>T, *LEPRg*.35592G>A, *LEPRg*.35657G>C and *LEPRg*.35856G>A were selected for differential expression analyses as representative of the four most informative groups of cosegregating polymorphisms (MAFs of 0.12, 0.22, 0.24 and 0.09, respectively). It is important to note that for all these promoter SNPs, one of the homozygous genotypes was detected in just one or none of the individuals; thereby these homozygous were discarded in further analyses.

Significant *LEPRglobal* expression differences were detected in liver and backfat conditional on *LEPRg*.34996C>T ($p<0.05$) and *LEPRg*.35592G>A ($p<0.05$) polymorphisms, however with opposite effects for the last SNP (Table 2). In addition, high *LEPRglobal* expression differences in liver were detected according to *LEPRg*.35856G>A, a SNP unidentified in previous studies. For this SNP, the GA genotype showed almost ten times higher expression than GG in liver. Regarding the specific *LEPR* isoforms, *LEPRb* resulted differentially expressed in hypothalamus according to *LEPR*g.34996C>T SNP, (Table 2), in the same sense that the observed for *LEPRglobal* in backfat and liver. Finally, significant *LEPRa* differential expression in liver and backfat were also detected for several polymorphisms. Expression differences of *LEPRa* in backfat tissue were found conditional on *LEPRg*.34996C>T and *LEPRg*.35592G>A SNPs, and in liver according to *LEPRg*.35856G>A, *LEPRg*.35592G>A and *LEPRg*.35657G>C (Table 2). As for *LEPRglobal*, the highest differences in hepatic gene expression were obtained for the *LEPRg*.35856G>A, where, same as before, the carriers of the GA genotype showed higher *LEPRa* expression than the homozygous GG. In order to validate this last result, due its large expression differences, eleven additional females were included into the analyses (6 AG and 5 GG animals). Note that only one animal of the whole pedigree was carrier of the AA genotype and therefore this genotype effect could not be tested. The addition of more samples allowed the validation of the expression differences detected revealing significant *LEPRa* differential expression ($p<0.001$) with a fold change of 15.77 between GA and GG genotypes (Table 2).

LEP Expression According to *LEP*g.1387C>T and *LEP* Promoter SNPs

The results from the Haploview software showed that all SNPs located in the *LEP* promoter are completely linked (MAF = 0.30), conversely *LEP*g.1387C>T is not fully linked (mean $r^2 = 0.5$) (Figure 3B). No significant expression differences could be detected for *LEP* gene in the analyzed tissues, according to the previously identified *LEP*g.1387C>T SNP, and neither to the promoter polymorphisms (Tables S4 and S5).

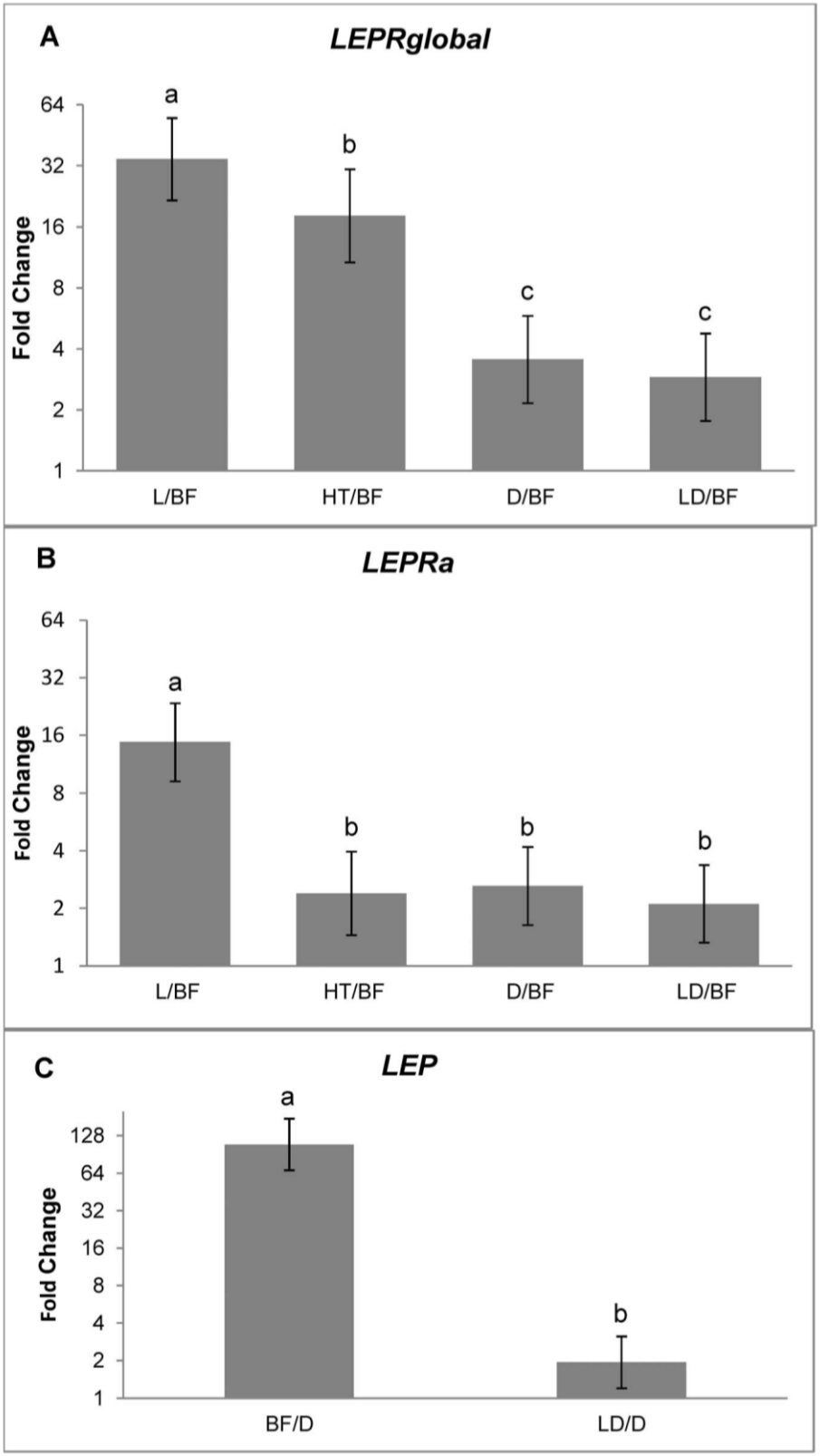


Figure 2. Pairwise comparison of gene expression values observed across the five tissues tested. A) *LEPRglobal* of all tissues related to backfat (the tissue that showed the lowest *LEPRglobal* expression). **B)** *LEPRa* of all tissues related to backfat (the tissue that showed the lowest *LEPRa* expression) **C)** *LEP* of all tissues related to diaphragm (the tissue that showed the lowest *LEP* expression). L: liver; BF: backfat; HT: hypothalamus; D: diaphragm; LD: *Longissimus dorsi*. Levels not connected by the same letter are significantly different ($p<0.05$).
doi:10.1371/journal.pone.0066398.g002

Table 1. Differential expression of *LEPR* transcripts conditional on *LEPRc.*1987C>T SNP genotypes.

Isoform	Tissue	Comparison	FC	Estimator	SE	95% CI	p-value
LEPRglobal	Backfat	CC-TT	5.447	−2.446	0.880	1.648–18.002	0.006
		CT-TT	2.984	−1.577	0.714	1.131–7.873	0.028
		CC-CT	1.826	−0.869	0.675	0.730–4.565	0.200
	Liver	CC-TT	7.783	−2.960	0.713	2.956–20.494	<.0001
		CT-TT	4.391	−2.135	0.664	1.782–10.819	0.002
		CC-CT	1.773	−0.826	0.496	0.903–3.478	0.098
LEPRb	Hypothalamus	CC-TT	2.715	−1.441	0.690	1.064–6.930	0.046
		CT-TT	2.794	−1.482	0.701	1.077–7.244	0.043
		CC-CT	0.970	0.049	0.605	0.430–2.210	0.946
LEPRa	Backfat	CC-TT	5.086	−2.347	0.819	1.673–15.464	0.005
		CT-TT	3.031	−1.600	0.669	1.222–7.518	0.018
		CC-CT	1.678	−0.747	0.629	0.714–3.943	0.237
	Hypothalamus	CC-TT	0.418	1.258	0.694	0.163–1.073	0.072
		CT-TT	0.313	1.678	0.708	0.119–0.818	0.019
		CC-CT	1.338	−0.420	0.611	0.583–3.070	0.493
	Liver	CC-TT	6.298	−2.655	0.775	2.197–18.061	0.001
		CT-TT	2.870	−1.521	0.671	1.153–7.144	0.025
		CC-CT	2.194	−1.134	0.575	1.005–4.791	0.050

FC: fold change; SE: standard error; CI: confidence interval.
doi:10.1371/journal.pone.0066398.t001

Discussion

The leptin hormone (*LEP*) interacts with its receptor (*LEPR*) to regulate food intake and energy expenditure influencing thereby important traits as growth and fatness [1]. Although *LEPR* gene presents several isoforms in other species, the *LEPRb* isoform is the best known and characterized. To our knowledge, none porcine *LEPR* isoform has been described rather than *LEPRb*. In the present study, a porcine short *LEPR* isoform has been described for the first time (*LEPRa*), homologous to human *LEPRa*. Compared with the porcine *LEPRb*, this novel isoform differs in the exon 20, in agreement with previous studies of *LEPR* isoforms in other species [4,5]. The *LEPRa* and *LEPRb* isoforms identified in the present study share identical extracellular and transmembrane N-terminus regions, but differ in the intracellular C-terminus. The predicted FN3 and transmembrane domains are located in the shared region. However, the longer intracellular region of the *LEPRb* contains the tyrosine residues responsible for intracellular signal transduction at hypothalamic level. An isoform lacking these residues would not be able to transduce the *LEP* signal by the same mechanism than *LEPRb* [37].

To get further insights on the potential functions of these *LEPR* isoforms, their expression patterns across five different porcine tissues have been investigated. Previously, a selection of control genes for expression data normalization was conducted. The most critical step in measuring gene expression is an accurate

Table 2. Differential expression of *LEPR* transcripts conditional on genotypes of different *LEPR* promoter SNPs.

SNP	Isoform	Tissue	Comparison	FC	Estimator	SE	95% CI	p-value
LEPR34996C>T	*LEPRglobal*	Backfat	CT-TT	0.356	1.488	0.503	0.180–0.706	0.003
		Liver	CT-TT	0.306	1.707	0.497	0.156–0.602	0.001
	LEPRb	Hypothalamus	CT-TT	0.418	−1.258	0.577	0.917–0.191	0.038
	LEPRa	Backfat	CT-TT	0.333	1.587	0.537	0.161–0.690	0.003
LEPR35856G>A	*LEPRglobal*	Liver	GA-GG	9.875	−3.304	0.456	5.316–18.343	<.0001
	LEPRa	Liver	GA-GG	15.770	−3.979	0.595	7.027–35.391	<.0001
LEPR35592G>A	*LEPRglobal*	Backfat	GA-AA	0.381	−1.391	0.455	0.708–0.205	0.003
		Liver	GA-AA	1.873	0.906	0.435	3.384–1.037	0.039
	LEPRa	Backfat	GA-AA	0.471	−1.087	0.497	0.925–0.239	0.030
		Liver	GA-AA	2.697	1.431	0.497	5.299–1.372	0.005
LEPR35657G>C	*LEPRa*	Liver	GC-CC	2.294	1.198	0.495	4.492–1.171	0.016

FC: fold change; SE: standard error; CI: confidence interval.
doi:10.1371/journal.pone.0066398.t002

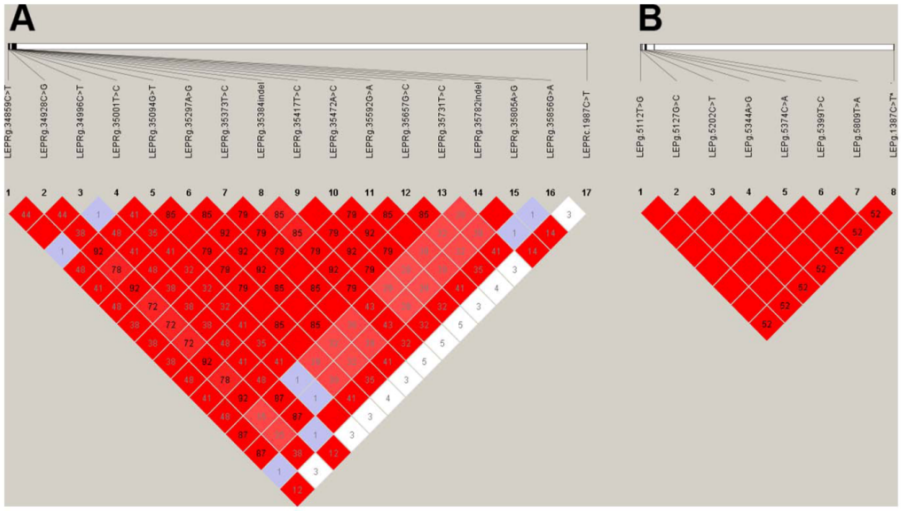


Figure 3. Linkage disequilibrium analysis among the SNPs detected for *LEPR* (A) and *LEP* (B) genes. The numbers in the boxes represent the linkage disequilibrium estimated with the r^2 value. Those boxes without a number indicate a complete linkage.
doi:10.1371/journal.pone.0066398.g003

normalization using suitable reference control genes. Even though numerous studies have evaluated different reference genes in several species, most of them have been directed towards specific types of tissues [30,33]. The difficulty increases notably when searching for reference genes exhibiting constant RNA transcription across tissues [38,39]. Even more when these tissues display great histological and metabolic differences, such as the five tissues analyzed. In the present study, the stability of five reference genes has been analyzed. According to the geNorm analyses and taking into account also PCR efficiencies of the five reference genes (*GADPH*, *TBP*, *TOP2B*, *B2M* and *ACTB*), *ACTB* and *B2M* genes were selected for normalizing the RT-PCR data.

The choice of a linear mixed model framework for estimating differential expression was justified by the complex design involving diverse experimental factors and biological and technical sources of variation. Simulation studies performed by Steibel et al. [36] showed that models fitting random sample effects (as D_{im}) and random interaction between sample and gene factors (as B_{gim}) provided better type I error rates and confidence intervals coverage than other alternative models of RT-PCR data analysis. The usefulness of this type of models in different experimental situations has been confirmed in recent studies [40,41].

The *LEPRb* isoform was quantifiable only in hypothalamus, previously reported in the same animal material by Ovilo et al. [15]. Its expression was unquantifiable in liver, fat and muscle tissues. This result was expected as it is known that *LEPRb* isoform acts mainly on hypothalamic nuclei, and agrees with previous studies in other species [4,5]. Even though, some authors reported the quantification of *LEPRb* expression in peripheral tissues in pigs [42,43], they do not report isoform differences and it is likely that the expression measured corresponded to a mixture of isoforms, instead of specifically the long one [44,45]. Moreover, this mixture of isoforms would mainly correspond to *LEPR* short isoforms, as it is known that *LEPRb* is much less abundant than the short forms in peripheral tissues [46].

Previous studies in other species reported that the short *LEPR* isoforms are widely expressed across all tissues [8,44]. Our results support this idea and show a significantly higher expression level of the porcine *LEPRa* isoform in liver compared to the remaining tissues. Notably, this result contrasts with the results obtained for the *LEPRb* isoform. The fact that both isoforms have such a different expression pattern suggests that they may develop different roles on the LEP-LEPR axis.

Although liver is the tissue responsible for the handling and degradation of several hormones, it is not likely that the high expression of the *LEPRa* in the liver is related to leptin clearance. Several studies performed in human [47] and rodents [48,49] concluded that the kidney is the main tissue of leptin clearance. Still, some authors have tried to find evidences of leptin degradation in other tissues like liver or spleen in human. Garibotto et al [50] supported the hypothesis that splanchnic organs (liver, spleen and small intestine) contribute in a significant way to leptin clearance when renal metabolic activity and function decline. In contrast, Jensen et al. [51] stated that even if small amounts of leptin are cleared by splanchnic organs in humans; they are minimal compared with kidney leptin clearance. A peripheral specific action of leptin in the liver could be a plausible explanation for this high *LEPRa* expression. Some studies proposed a role of leptin in liver cholesterol metabolism, downregulating cholesterol biosynthesis, upregulating cholesterol catabolism and decreasing plasma levels of very low-density lipoprotein [52]. *In vitro* and *in vivo* evidences proposed leptin peripheral roles in modulating liver lipoprotein receptor levels to ensure efficient lipid removal following a meal and contributing to the dynamics of lipid distribution and utilization [53]. Within the frame of these leptin peripheral roles, *LEPRa* would be necessary in the liver to detect and respond to different amounts of leptin triggering the specific downstream actions.

The performed *LEPRglobal* expression measures allowed the estimation of the relation among total amount of *LEPR* isoforms and the specific *LEPRa* and *LEPRb*. High or moderate correlations were detected between *LEPRa* and *LEPRglobal* expression across tissues except in the hypothalamus. Actually, the correlation between both measures was 0.87 and 0.90 in backfat and liver respectively ($p<0.0001$), and 0.4 in both muscles ($p<0.05$). The high correlation found in backfat and liver suggest that *LEPRa* is the predominant isoform in these tissues, or if there are other short isoforms, they would show low expression or are regulated in the same way. The results found in muscles suggest that there are other isoforms, apart from *LEPRa* expressing in both muscles. As mentioned before, no significant correlation was found between *LEPRa* and *LEPRglobal* in hypothalamus (0.321; $p>0.1$), but it

should be taken into account that *LEPRb* is almost exclusively expressed in this tissue and included in the *LEPRglobal* measure, but not in *LEPRa* measure. Even more, *LEPRa* and *LEPRb* isoforms measures do not appear significantly correlated (−0.27; $p>0.1$), suggesting an isoform specific regulation at the hypothalamic level. A plausible explanation of the results obtained could be an autorregulatory mechanism, as both isoforms seem to develop different functions in hypothalamus. Therefore, an increase of *LEPRb* in the hypothalamus would inhibit *LEPRa* expression in this tissue. These mechanisms of autoregulation among isoforms have been previously reported for genes such as *Pax6* [54] and *ZFHX1A* [55].

The only *LEP* transcript (expands from exon 2 to exon 3) showed a significant higher expression on backfat compared to the other tested tissues. This highly specific backfat *LEP* expression was expected as it is known that *LEP* is mainly synthesized by white adipocytes. However, the detection of *LEP* expression in both muscles, even at low levels compared to backfat, gives more evidence of the potential peripheral roles of this hormone.

In order to get other insights on *LEP* and *LEPR* transcriptional regulation, gene expression differences conditional on several polymorphisms have been investigated. Our previous study revealed that *LEPRb* hypothalamic expression is conditioned by *LEPR*c.1987C>T polymorphism, the animals carrying the T allele showed a lower expression than carriers of the C allele [15]. This result is confirmed in the present study using a more suitable statistical approach. A similar effect of this SNP was also found on *LEPRa* expression in the liver and backfat. Interestingly, an opposite effect of *LEPR*c.1987C>T on *LEPRa* hypothalamic expression was found, the TT carriers showed higher expression that CT. In agreement with these conflicting result, no effect of *LEPRc*.1987C>T was found when analyzing the *LEPRglobal* expression in hypothalamus. These results support the hypothesis of isoform specific regulation at the hypothalamic level. The *LEPR* promoter was previously described by Lee et al. [27], but differential expression analyses have never been performed before conditional on SNP genotypes located in this region. Here, 16 polymorphisms have been identified in the *LEPR* promoter region, 14 of which are located within potential TF binding sites. The analysis of these polymorphisms revealed several significant differential expressions in hypothalamus, liver and backfat. However, it should be taken into account that for all these SNPs only two segregating genotypes appeared (one of the homozygous was found in one or none individuals), and most of them are highly linked (Figure 3A). Additionally, multiple statistic tests have been performed, therefore some of these results could be significant by random and should be taken with caution. Nonetheless, a high consistency was found between *LEPRa* and *LEPRglobal* expression differences according to the promoter SNPs analyzed within tissue (liver and backfat). The results obtained for the novel *LEPR*g.35856G>A promoter SNP on both *LEPRa* and *LEPRglobal* in liver were especially relevant, due to the high expression differences detected. This SNP is representative of two SNPs, *LEPRg*.35856G>A and *LEPRg*.35001T>C, that are predicted to fall within TF binding sites (Table S2). More specifically, the *LEPR*g.35856G>A polymorphism falls within a predicted target for the *CREB* transcription factor ($p<0.05$; Table S2), which is implicated in expression regulation of several genes affecting appetite [56]. The results obtained for *LEPR* expression analyses strongly suggest a tissue and isoform specific regulation focused mainly on the effects of the exonic *LEPRc*.1987C>T SNP on the regulation of both *LEPRb* and *LEPRa* expression in different tissues, but with opposite effects on both isoforms in hypothalamus, and the promoter *LEPR*g.35856G>A SNP that would specifically regulate *LEPRa* expression in hepatic tissue.

Finally, non-differential expression of *LEP* conditional on any of the analyzed polymorphisms could be detected. These results agree with the study of Stachowiak et at. [26], which did not find association of the *LEP* expression levels in subcutaneous fat with the polymorphisms identified in the *LEP* promoter region. Although Liu et al. [57] reported differential expression of the *LEP* gene in subcutaneous fat tissue due to a polymorphism detected outside the functional promoter region; our results seem to indicate that gene expression regulation is likely affected by other mechanisms rather than gene sequence variants. Diet and changes in nutrient availability result in rapid alterations in *LEP* autoregulation mechanisms [58]. In addition, the adipocyte volume and several metabolites such as insulin, glucose, glucocorticoids or cytokines have been shown to affect *LEP* expression [59].

An overview of the results obtained in the present study suggests an implication of both molecules (LEP and LEPR) on liver metabolism. In addition, the specific expression patterns of the *LEPR* isoforms highlight the idea that they are involved in different biological pathways, supporting previous studies that suggest peripheral roles of porcine LEP and LEPR. In fact, previous studies reported effects of *LEP* and *LEPR* genes polymorphisms on fatty acid composition in backfat and muscle suggesting local effect of LEP-LEPR on fatty acids metabolism, specifically on oleic fatty acid level [16]. Similarly, Galve et al. [20] suggested a local effect of LEP and LEPR in skeletal muscle to explain the changes in the proportions of saturated and unsaturated fatty acids observed in different muscle tissues. The present study has allowed a transcriptional characterization of *LEP* and *LEPRa* and *LEPRb* isoforms on a range of tissues. Complementary functional studies would be required in order to determine the peripheral *LEP* and *LEPR* gene function and regulation.

Supporting Information

Figure S1 Average expression stability values of the reference genes tested according to the geNorm analysis. Those genes with stability values lower than 1.5 could be considered as appropriate reference genes.
(TIF)

Table S1 Primer pairs and PCR conditions for expression analyses, *LEPR* isoform detection and promoters sequencing.
(DOC)

Table S2 Description of the polymorphisms detected in the promoter regions of *LEP* and *LEPR* genes.
(DOCX)

Table S3 Pairwise comparison of gene expression values observed across the five tissues tested for *LEPRglobal, LEPRa* and *LEP* isoforms.
(DOCX)

Table S4 Differential *LEP* expression conditional on *LEPg.1387T>C* genotype.
(DOCX)

Table S5 Differential *LEP* expression conditional on *LEP* promoter SNPs.
(DOCX)

Acknowledgments

We want to thank Rita Benítez, Yolanda Nuñez and Fabián García for their technical support.

Author Contributions

Conceived and designed the experiments: MCR CO LS AIF. Performed the experiments: DPM CB AIF. Analyzed the data: DPM AF LS AIF. Contributed reagents/materials/analysis tools: MCR JMF JLN LS. Wrote the paper: DPM AIF.

References

1. Barb CR, Hausman GJ, Houseknecht KL (2001) Biology of leptin in the pig. Domest Anim Endocrinol 21(4): 297–317.
2. Houseknecht KL, Baile CA, Matteri RL, Spurlock ME (1998) The biology of leptin: a review. Journal of Animal Science 76: 1405–20.
3. Margetic S, Gazzola C, Pegg GG, Hill RS (2002) Leptin: a review of its peripheral actions and interactions. International Journal of Obesity 26: 1407–1433.
4. Lee GH, Proenca R, Montez JM, Carroll KM, Darvishzadeh JG, et al. (1996) Abnormal splicing of the leptin receptor in diabetic mice. Nature 379: 632–635.
5. Bacart J, Leloire A, Levoye A, Froguel P, Jockers R, et al. (2010) Evidence for leptin receptor isoforms heterodimerization at the cell surface. FEBS Lett 584(11): 2213–2217.
6. Mercer JG, Hoggard N, Williams LM, Lawrence CB, Hannah LT, et al. (1996) Localization of leptin receptor mRNA and the long form splice variant (Ob-Rb) in mouse hypothalamus and adjacent brain regions by in situ hybridization. FEBS Lett 387: 113–116.
7. Schwartz MW, Peskind E, Raskind M, Boyko EJ, Porte D Jr (1996) Cerebrospinal fluid leptin levels: relationship to plasma levels and to adiposity in humans. Nat Med 2: 589–593.
8. Fei H, Okano HJ, Li C, Lee GH, Zhao C, et al. (1997) Anatomic localization of alternatively spliced leptin receptors (OBR) in mouse brain and other tissues. Proc Natl Acad Sci U S A 94: 7001–7005.
9. Elmquist JK, Børbæk C, Ahima RS, Flier JS, Saper CB (1998) Distributions of leptin receptor mRNA isoforms in the rat brain. J Comp Neurol 395: 535–547.
10. Uotani S, Bjørbaek C, Trornoe J, Flier JS (1999) Functional properties of leptin receptor isoforms: internalization and degradation of leptin and ligand-induced receptor downregulation. Diabetes 48(2): 279–286.
11. Tartaglia LA, Dembski M, Weng X, Deng N, Culpepper J, et al. (1995) Identification and expression cloning of a leptin receptor, OB-R. Cell 83: 1263–1271.
12. Wylie ARG (2011) Leptin in farm animals: where are we and where can we go? Animal 5: 246–267.
13. Switonski M, Stachowiak M, Cieslak J, Bortz M, Grzes M (2010) Genetics of fat tissue accumulation in pigs: a comparative approach. Journal of Applied Genetics 51: 153–168.
14. Óvilo C, Fernández A, Noguera JL, Barragán C, Letón R, et al. (2005) Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. Genet Res 85: 57–67.
15. Óvilo C, Fernández A, Fernández AI, Folch JM, Varona L, et al. (2010) Hypothalamic expression of porcine *leptin receptor* (*LEPR*), *neuropeptide Y* (*NPY*), and *cocaine- and amphetamine-regulated transcript* (*CART*) genes is influenced by *LEPR* genotype. Mammalian Genome 21: 583–591.
16. Pérez-Montarelo D, Fernández A, Folch JM, Pena RN, Óvilo C, et al. (2012) Joint effects of porcine *leptin* and *leptin receptor* polymorphisms on productivity and quality traits. Animal Genetics 43(6): 805–809.
17. Muñoz G, Óvilo C, Silió L, Tomás A, Noguera JL, et al. (2009) Single- and joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on Sus scrofa chromosome 6 and leptin receptor effects on fatness and growth traits. J Anim Sci 87(2): 459–468.
18. Muñoz G, Alcazar E, Fernández A, Barragán C, Carrasco A, et al. (2011) Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc x Iberian crossbred pigs. Meat Sci 88(1): 163–173.
19. Rodríguez MC, Fernández A, Carrasco C, García A, Gómez E, et al. (2010) Effect of LEPRc2002C>T SNP on feed intake and growth in heavy Duroc x Iberian crossbred pigs. Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production: 1–6 August 2010: Leipzig.
20. Galve A, Burgos C, Silió L, Varona L, Rodríguez C, et al. (2012). The effects of leptin receptor (LEPR) and melanocortin-4 receptor (MC4R) polymorphisms on fat content, fat distribution and fat composition in a Duroc x Landrace/Large White cross. Livestock Science 145: 145–152.
21. Ovilo C, Pérez-Enciso M, Barragán C, Clop A, Rodríguez C, et al. (2000) A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6. Mamm Genome 11(4): 344–346.
22. Artimo P, Jonnalegedda M, Arnold K, Baratin D, Csardi G, et al. (2012) ExPASy: SIB bioinformatics resource portal. Nucl Acids Res 40(W1): W597–W603.
23. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucl Acids Res 40(D1): D302–D305.
24. Prestridge DS (1995) Predicting Pol II promoter sequences using transcription factor binding sites. J Mol Biol 249: 923–932.
25. Scherf M, Klingenhoff A, Werner T (2000) Highly Specific Localization of Promoter Regions in Large Genomic Sequences by PromoterInspector: A Novel Context Analysis Approach. J Mol Biol 297(3): 599–606.
26. Stachowiak M, Mackowski M, Madeja Z, Szydlowski M, Buszka A, et al. (2007) Polymorphism of the porcine leptin gene promoter and analysis of its association with gene expression and fatnesss traits. Biochem Genet 45: 245–253.
27. Lee KT, Hwang H, Kang KS, Park EW, Kin JH, et al. (2008) Genomic structure of porcine leptin receptor gene (*LEPR*)/LEPR overlapping transcript gene (*LEPROT*) and SNP discovery, association study. Congress of the International Society of Animal Genetics 2008.
28. Ghosh D (2000) Object-oriented transcription factors database (ooTFD). Nucleic Acids Res 28(1): 308–310.
29. Luu-The V, Paquet N, Calvo E, Cumps J (2005) Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. BioTechniques 38: 287–293.
30. Gu Y, Li M, Zhang K, Chen L, Jiang AA, et al. (2011) Identification of suitable endogenous control microRNA genes in normal pig tissues. Anim Sci J 82(6): 722–728.
31. Svobodová K, Bílek K, Knoll A (2008) Verification of reference genes for relative quantification of gene expression by real-time reverse transcription PCR in the pig. J Appl Genet 49(3): 263–265.
32. Kuijk EW, du Puy L, van Tol HTA, Haagsman HP, Colenbrander B, et al. (2007) Validation of reference genes for quantitative RT-PCR studies in porcine oocytes and preimplantation embryos. BMC Developmental Biology 7: 58–66.
33. Erkens T, Van Poucke M, Vandesompele J, Goossens K, Van Zeveren A, et al. (2006) Development of a new set of reference genes for normalization of real-time RT-PCR data of porcine backfat and *Longissimus dorsi* muscle, and evaluation with *PPARGC1A*. BMC Biotechnology 6: 41–48.
34. Vandesompele J, De Preter K, Pattyn F, Pappe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7).
35. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–265.
36. Steibel JP, Poletto R, Coussens PM, Rosa GJM (2009) A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. Genomics 94: 146–152.
37. Ceddia RB (2005) Direct metabolic regulation in skeletal muscle and fat tissue by leptin: implications for glucose and fatty acids homeostasis. International Journal of Obesity 29: 1175–1183.
38. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, et al. (2004) Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 313(4): 856–862.
39. Nygard AB, Jørgensen CB, Cirera S, Fredholm M (2007) Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. BMC Molecular Biology 8: 67.
40. Brown DE, Dechow CD, Liu WS, Harvatine KJ, Ott TL (2012) Hot topic: Association of telomere length with age, herd, and culling in lactating Holsteins. J Dairy Sci 95: 6384–6387.
41. Rezamand P, Watts JS, Hunt KM, Bradford BJ, Mamedova LK, et al. (2012) Bovine hepatic and adipose retinol-binding protein gene expression and relationship with tumor necrosis factor-alpha. J Dairy Sci 95(12): 7097–7104.
42. Tyra M, Ropka-Molik K, Eckert R, Piórkowska K, Oczkowicz M (2010) H-FABP and LEPR gene expression profile in skeletal muscles and liver during ontogenesis in various breeds of pigs. Domestic Animal Endocrinology 40: 147–154.
43. Xiaoxiang HU, Rujuan DAI, Ning LI, Changxin WU (2001) Expression detection and partial cloning of porcine leptin receptor (OBR) gene. Chinese Sci Bull 46: 396–401.
44. Bogacka I, Przala J, Siawrys G, Kaminski T, Smolinska N (2006) The expression of short form of leptin receptor gene during early pregnancy in the pig examined by quantitative real time RT-PCR. J Physiol Pharmacol 57(3): 479–489.
45. Kaminski T, Smolinska N, Gajewska A, Siawrys G, Okrasa S, et al. (2006) Leptin and long form of leptin receptor genes expression in the hypothalamus and pituitary during the luteal phase and early pregnancy in pigs. J Physiol Pharmacol 57(1): 95–108.
46. Tartaglia LA (1997) The leptin receptor. The Journal of Biological Chemistry 7: 6093–6096.
47. Meyer C, Robson D, Rackovsky N, Nadkarni V, Gerich J (1997) Role of the kidney in human leptin metabolism. Am J Physiol 273: 903–907.
48. Cumin F, Baum HP, Levens N (1996) Leptin is cleared from the circulation primarily by the kidney. Int J Obes 20: 1120–1126.
49. Zeng J, Patterson BW, Klein S, Martin DR, Dagogo-Jack S, et al. (1997) Whole body leptin kinetics and renal metabolism in vivo. Am J Physiol 273: 1102–1106.
50. Garibotto G, Russo R, Franceschini R, Robaudo C, Saffioti S, et al. (1998) Inter-organ leptin exchange in humans. Biochem and Biophys Res 247: 504–509.

51. Jensen MD, Møller N, Nair KS, Eisenberg P, Landt M, et al. (1999) Regional leptin kinetics in humans. Am J Clin Nutr 69: 18–21.
52. VanPatten S, Ranginani N, Shefer S, Nguyen LB, Rossetti L, et al. (2001) Impaired biliary lipid secretion in obese Zucker rats: leptin promotes hepatic cholesterol clearance. Am J Physiol Gastrointest Liver Physiol 281: 393–404.
53. Stenger C, Hanse M, Pratte D, Mbala ML, Akbar S, et al. (2010) Up-regulation of hepatic lipolysis stimulated lipoprotein receptor by leptin: a potential lever for controlling lipid clearance during the postprandial phase. FASEB J 24: 4218–4228.
54. Pinson J, Simpson TI, Mason JO, Price DJ (2006) Positive autoregulation of the transcription factor Pax6 in response to increased levels of either of its major isoforms, Pax6 or Pax6(5a), in cultured cells. BMC Developmental Biology 6: 25–34.
55. Manavella PA, Roqueiro G, Darling DS, Cabanillas AM (2007) The ZFHX1A gene is differentially autoregulated by its isoforms. Biochem Biophys Res Commun 360(3): 621–626.
56. Yadav VK, Oury F, Tanaka KF, Thomas T, Wang Y, et al. (2011) Leptin-dependent serotonin control of appetite: temporal specificity, transcriptional regulation, and therapeutic implications. J Exp Med 208(1): 41–52.
57. Liu D, Hu Y, Yang X, Liu Y, Wei S, et al. (2011) Identification and genetic effects of a novel polymorphism in the distal promoter region of porcine leptin gene. Mol Biol Rep 38(3): 2051–2057.
58. Wang J, Liu R, Liu LS, Chowdhury R, Barzilail N, et al. (1999) The effect of leptin on *LEP* expression is tissue-specific and nutritionally regulated. Nature Medicine 5: 895–899.
59. Wrann CD, Rosen ED (2012) New insights into adipocyte-specific leptin gene expression. Adipocyte 1(3): 168–172.

DISCUSIÓN

En los últimos años se ha producido un gran avance en lo que a técnicas de análisis genómico se refiere. Existe una gran diferencia entre los primeros estudios de detección de QTLs llevados a cabo con un número reducido de microsatélites y la información generada por miles de SNPs contenidos en los chips de genotipado masivo que se utilizan actualmente, que permite refinar mucho más los intervalos de confianza de los QTLs identificados, facilitando la selección de genes candidato posicionales, así como identificar nuevos QTLs en regiones no cubiertas anteriormente por marcadores informativos. Igualmente, los estudios de expresión génica han sido objeto de muy importantes mejoras que permiten pasar del estudio individual de cada gen al análisis global del transcriptoma, primero centrándose solo en regiones del genoma conocidas, en el caso de los microarrays y después analizando todas las regiones del genoma expresadas mediante la tecnología de RNA-seq.

El objetivo de la presente tesis doctoral ha sido caracterizar la base genética del crecimiento y la deposición grasa en el cerdo, utilizando tanto técnicas de vanguardia de análisis masivo y global del genoma, como estrategias más tradicionales en los casos que así lo han requerido. A través de las distintas aproximaciones utilizadas se han obtenido resultados relevantes sobre genes y polimorfismos concretos relacionados con estos caracteres.

El estudio clásico de detección de QTLs utilizando información de genotipado masivo de SNPs ha permitido, en primer lugar, confirmar dos de los QTLs previamente identificados en la población IBMAP (localizados en SSC4 y SSC6) con intervalos de confianza que se han reducido de 12 y 18 cM a solo 7 y 2 cM respectivamente, y en segundo lugar, identificar nuevas regiones del genoma asociadas con caracteres de crecimiento como el peso de la canal y el peso a 150 días, y de deposición grasa como el espesor de tocino y el contenido de grasa intramuscular. Se ha utilizado una estrategia muy conservadora a la hora de seleccionar los SNPs a utilizar en el análisis de ligamiento, seleccionando sólo 8.417 SNPs de los más de 60.000 contenidos en el chip. Se han aplicado criterios estrictos de calidad y se ha tenido en cuenta la posición e informatividad de los marcadores utilizados según el estudio de mapeo de ligamiento realizado en este mismo material por Muñoz et al., 2011. Estos análisis, que se realizaron utilizando l

conjuntamente las generaciones F3 y los dos retrocruces de la población IBMAP o separándola de acuerdo a la generación machos utilizados para su obtención (individuos F2 en el caso de las generaciones F3 y uno de los retrocruces e individuos F1 en el caso del segundo retrocruce), mostraron las diferencias entre los genotipos para los QTL esperados y el muestreo aleatorio de los alelos de los QTLs en los individuos F1 y F2. Varios de los QTLs identificados, como el del cromosoma 11 para espesor de tocino dorsal a 75kg, pudieron ser detectados en el conjunto de la población, mientras que otros fueron únicamente detectados en uno de los dos grupos de animales. En total se han detectado nueve QTLs nuevos asociados a los caracteres de interés en los cromosomas SSC 1, 2, 5, 9, 11, 13, 14 y 17, con intervalos de confianza casi siempre inferiores a 5cM. La reducción de tamaño de las regiones identificadas facilitó la identificación de posibles genes candidatos posicionales que además son candidatos funcionales. Mediante esta estrategia se han identificado varios nuevos genes candidatos que sería interesante estudiar en mayor profundidad por su relación con los caracteres de interés. Así se han propuesto los siguientes genes candidatos: *PTPRD, VCAM1, AGL, ADIPOR2, VDR, MYOG, FASLG, NCK1, CYP2E1* e *ID1*.

El estudio del transcriptoma completo mediante RNAseq ha permitido la identificación de otros genes relacionados con crecimiento y deposición grasa usando una estrategia totalmente diferente, complementando así la identificación de genes potencialmente asociados a este tipo de caracteres que no han sido identificados en la anterior aproximación. Se ha llevado a cabo por primera vez el estudio del transcriptoma completo mediante RNA-seq del hipotálamo porcino, evidenciando la mayor complejidad del transcriptoma hipotalámico en comparación con otros tejidos y la gran relevancia de este tejido en la regulación de los caracteres de interés. Además este estudio ha puesto de manifiesto, una vez más, la necesidad de continuar los esfuerzos para mejorar la anotación del genoma porcino disponible en la actualidad, ya que siguen existiendo un gran número de regiones no anotadas y errores en la anotación como muestran el elevado número de tránscritos identificados como intergénicos o que mapean en regiones teóricamente intrónicas. El estudio de expresión diferencial ha permitido identificar tanto genes anotados como nuevas isoformas diferencialmente expresados entre los animales divergentes para los caracteres de interés, que han

sido validados con otras tecnologías, considerando como genes candidatos más potentes aquellos que se localizan dentro de las regiones QTL identificadas en el estudio anterior. Estos genes resultan más interesantes debido a que, además de localizarse en las regiones QTL y estar diferencialmente expresados en los grupos de cerdos divergentes para estos caracteres, tienen funciones asociadas a los caracteres de interés o se trata de potenciales reguladores de redes o rutas involucradas en estos caracteres. De la lista de genes e isoformas diferencialmente expresadas identificados mediante esta estrategia cabe resaltar los genes *EGR-1* e *IRF1*, por ser tanto candidatos posicionales, localizados en el QTL SSC2, como funcionales. El gen *IRF1* codifica un factor de transcripción involucrado en la regulación de la hormona de crecimiento (Hu et al. 2012). El *EGR-1* codifica así mismo un factor de trancripción que juega un importante papel en la regulación de la ingesta de alimentos y de la duración y cantidad de la ingesta (de Lartigue et al. 2010). Junto con ellos, aunque no aparezca diferencialmente expresado en este estudio, cabe destacar también el gen *NR3C1*, ya que se trata de un factor de transcripción que regula la expresión de 5 genes identificados como diferencialmente expresados en el análisis. Este gene mapea en el intervalo de confianza del mismo QTL SSC2 y codifica para el receptor de glucocorticoides, asociado previamente con la obesidad y desórdenes alimentarios en humanos (Cellini et al. 2010). En base a los resultados obtenidos, estos tres factores de transcripción parecen ser importantes reguladores de genes relacionados con el crecimiento y la deposición grasa y resultan interesantes candidatos para futuros estudios. Ambos trabajos nos han brindado la oportunidad de, analizando la información global, centrar el estudio en una serie de regiones genómicas y genes que resulten relevantes.

Sin embargo, en este punto y al menos por el momento, es necesario dejar de lado los análisis masivos y centrarse en genes concretos para analizarlos más en profundidad y confirmar el posible efecto de los mismos sobre los caracteres de interés.

A pesar de la gran cantidad de información generada mediante estas técnicas de análisis masivo, éstas también presentan algunas desventajas. Los posibles errores en el mapeo de genes con secuencias similares pueden generar resultados

incorrectos. Además, la necesidad de normalizar los datos y corregirlos, aunque sin duda se ha reducido mucho respecto a la tecnología de los microarrays de expresión, sigue existiendo y también puede producir sesgos en los resultados (McGettigan 2013). Otra de las desventajas que presentan en la actualidad estos análisis en porcino es que los errores antes mencionados en la anotación del genoma pueden llevar a errores en la interpretación de los resultados, además de que se pueda estar perdiendo mucha información en las regiones que no están adecuadamente anotadas. Un claro ejemplo de una región en la cual no se ha podido llevar a cabo un estudio mediante técnicas de análisis masivo sería el QTL identificado en el SSC6 y el estudio del gen *LEPR* como gen candidato a portar la mutación causal, que al no estar correctamente anotado en el genoma porcino no se puede estudiar mediante técnicas masivas y es necesario analizarlo de forma particular.

Debido al gran impacto tanto de la *LEP* como de su receptor (*LEPR*) en la señalización hipotalámica y los resultados previos de la mutación *LEPR*c.1987C>T, ambos genes han sido estudiados en profundidad usando diversas estrategias de análisis tanto estructurales como de expresión génica. La región codificante del gen *LEP* y parte de las regiones adyacentes han sido secuenciadas en los animales de la población IBMAP. Los resultados de asociación más relevantes se han obtenido con el polimorfismo *LEP*g.1387C>T que ha mostrado asociación significativa con caracteres de crecimiento, deposición grasa, conformación y composición de ácidos grasos, demostrando la relevancia de este gen sobre los caracteres de interés. Además, se han confirmado los efectos detectados anteriormente del polimorfismo *LEPR*c.1987C>T y se han identificado por primera vez en este material efectos de este SNP sobre la composición de ácidos grasos en grasa subcutánea. En ambos SNPs el alelo T parece aumentar el apetito y la deposición grasa, aumentando con ello también el contenido de ácidos grasos saturados y reduciendo los ácidos grasos mono y poliinsaturados.

Adicionalmente, se ha llevado a cabo un análisis de la diversidad nucleotídica y haplotípica de las regiones más polimórficas de los dos genes en varias razas porcinas, representativas de poblaciones europeas, asiáticas y jabalí. Se ha detectado una gran variabilidad haplotípica en la región analizada del gen *LEP* en

todas las poblaciones analizadas, incluyendo el cerdo Ibérico donde se han detectado 18 haplotipos distintos en los 12 individuos analizados. Estos datos nos han permitido confirmar la divergencia entre las poblaciones asiáticas y europeas así como, corroborar la localización de dos núcleos de domesticación en Asia, sugiriendo además cierta subdivisión en los genomas de origen europeo no detectada en estudios previos, aunque este resultado requiere análisis complementarios ya que podría deberse a la existencia de haplotipos recombinantes o a un efecto del muestreo. Por el contrario, los resultados obtenidos del análisis de la variabilidad nucleotídica y haplotipica de la región promotora del gen *LEPR* han mostrado una menor variabilidad de este gen en las poblaciones analizadas, especialmente en el caso de la raza Ibérica donde no se ha detectado ninguna variación pese al muestreo de individuos de las distintas variedades morfológicas de esta raza. En contraste con la alta variabilidad identificada en esta raza a nivel genómico en trabajos previos como de fenotipo en estudios previos, en este caso sólo se detecta un haplotipo en el gen *LEPR* en cerdo ibérico, lo que podría indicar un posible efecto de selección de este gen probablemente asociado a la singularidad del entorno ambiental (las dehesas del suroeste de la península Ibérica) en que tuvo lugar la conformación de esta raza, con una fuerte variación estacional de los recursos pastables (bellota y hierba) disponibles.

A pesar de la relevancia funcional de los genes *LEP* y *LEPR*, no se conocen totalmente sus patrones de expresión en ninguna especie. En el caso de la leptina sólo se ha identificado una isoforma en la mayoría de los mamíferos y, aunque se sabe que su expresión se da fundamentalmente en el tejido adiposo, se ha detectado expresión de este gen en otros tejidos sugiriendo la posibilidad de que lleve a cabo funciones periféricas no descritas (Margetic et al. 2002). Al menos se han descrito seis isoformas del gen *LEPR* en mamíferos (Figura 11), siendo la isoforma más larga (*LEPRb*) la que desarrolla la función de señalización hipotalámica descrita. Sin embargo, los patrones de expresión del resto de isoformas de este gen y su función no han sido descritos en ninguna especie hasta el momento.

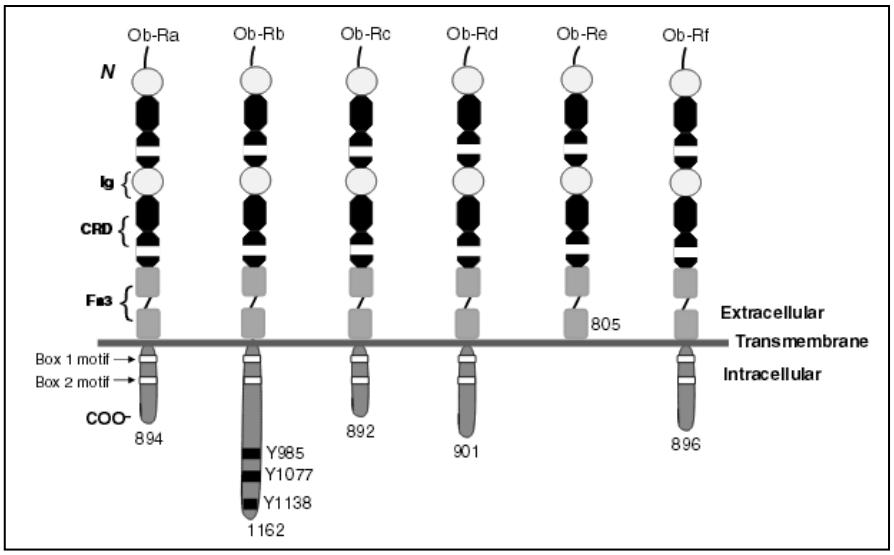


Figura 11. Isoformas del gen *LEPR* identificadas en humano y ratón (Ceddia 2005).

En la presente tesis doctoral, se ha identificado una isoforma corta del gen *LEPR* que no había sido previamente descrita en cerdo, y que es homóloga a la isoforma humana *LEPRa.* En humanos se ha detectado expresión de esta isoforma corta en la mayoría de los tejidos y aunque su función se desconoce, se ha sugerido que pueda estar relacionada con el transporte de la leptina a través del torrente sanguíneo o en su degradación. Los resultados del estudio de los patrones de expresión del gen *LEP* y de las dos isoformas identificadas del *LEPR* identificadas en porcino, *LEPRb* y *LEPRa*, indican que ambos genes presentan acciones periféricas, además de su conocida función en la ruta de señalización hipotalámica, y pueden estar involucradas en el metabolismo hepático. Además parecen sugerir que las diferentes isoformas del *LEPR* están implicadas en diferentes rutas biológicas ya que presentan distinto patrón de expresión y posiblemente distinta regulación transcripcional. El análisis de la región promotora del gen *LEPR* ha permitido identificar polimorfismos candidatos a regular los niveles de expresión de estos isoformas de manera tejido dependiente. Varios de los SNPs del promotor del *LEPR* han mostrado asociación con los niveles de expresión de algunas de las isoformas del *LEPR* en tejidos concretos, siendo la asociación más relevante aquella detectada en hígado para el polimorfismo *LEPR*g.35856G>A con la isoforma corta (*LEPRa)*, polimorfismo localizado en una potencial diana del factor de transcripción *CREB*,

asociado con la regulación de genes implicados en la regulación del apetito (Yadav et al. 2011).

Analizando los resultados obtenidos de forma global se puede decir que las diversas aproximaciones utilizadas en la presente tesis han permitido la identificación de genes y rutas génicas que parecen desarrollar importantes funciones en el crecimiento y la deposición grasa en porcino. De hecho, el análisis de interacción *in silico* (herramientas String y GeneMania) con los 34 genes más relevantes extraídos de los distintos trabajos presentados (los diez genes candidatos identificados en la detección de QTLs, los 22 genes destacados en el análisis del transcriptoma hipotalámico, y los genes *LEP* y *LEPR*) permite visualizar relaciones e interacciones, conocidas o predichas, entre 15 de ellos (Figura 12).

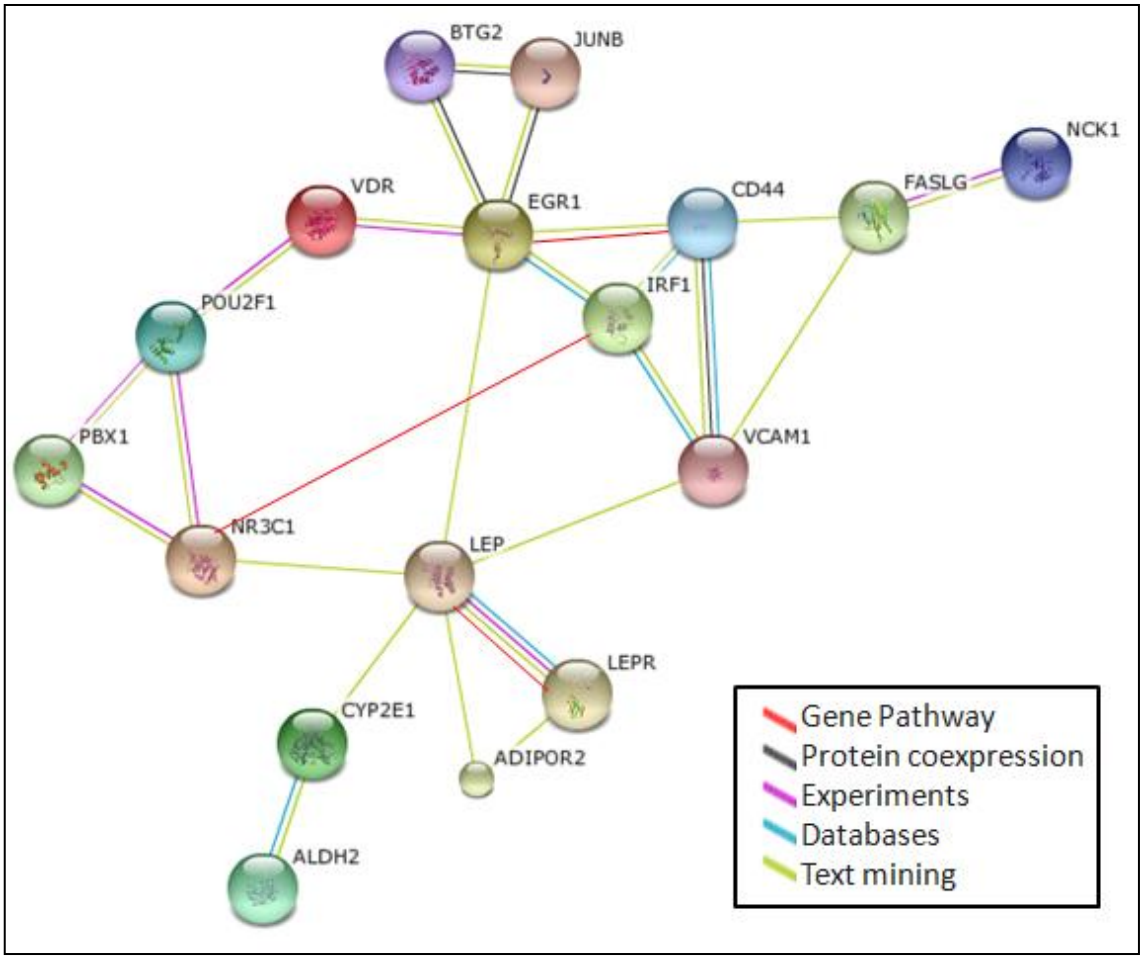


Figura 12. Red génica construida mediante el análisis de las interacciones entre las proteínas codificadas por los genes destacados en la presente tesis doctoral y potencialmente involucrados en crecimiento y deposición grasa en porcino (String y GeneMania).

La red obtenida pone de manifiesto una relación de la adiponectina *ADIPOR2*, identificado en el estudio de detección de QTLs como potencial candidato del QTL del cromosoma 5, tanto con la leptina como con su receptor sugiriendo una posible implicación de este gen en el eje *LEP-LEPR*. En la red queda patente además la interacción del producto del gen *EGR-1* con la leptina como se había mencionado anteriormente (de Lartigue et al. 2010). Sería interesante además analizar el posible papel del factor de transcripción *NR3C1* sobre el eje *LEP-LEPR*, dada la asociación de este gen con la obesidad y desórdenes alimentarios en humanos (Cellini et al. 2010) y su aparente asociación en la red generada. Cabe destacar de nuevo la localización central y el número de interacciones de las proteínas codificadas por los genes *EGR-1* e *IRF1,* conforme a la hipótesis de que se trate de importantes reguladores transcripcionales del control del crecimiento y la deposición grasa. Los productos de los genes *BTG2* y *JUNB* aparecen relacionados con el factor de transcripción *EGR-1*. Como se había sugerido ya en el análisis de RNA-seq, este factor podría considerarse regulador de ambos genes, por lo que todos los resultados parecen ratificar también el relevante papel de este trío de genes en los caracteres de interés. Desataca también en la red el triángulo formado por los factores de transcripción *PBX1*, *NR3C1* y *POU2F1* cuyo potencial papel regulador sobre los caracteres de interés también ha sido discutido anteriormente en el análisis de RNA-seq y que aparecen también conectados a los productos de otros genes relevantes. La relación de los productos de los genes *CD44*, recientemente asociado a la obesidad y la resistencia a insulina (Kang et al. 2013), y *VCAM1*, que contiene un SNP asociado al espesor del tocino dorsal en un trabajo previo (Fontanesi et al. 2012) y localizado dentro del QTL SSC4, con los productos génicos de *EGR-1* e *IRF1,* asevera la implicación de estos dos genes (*CD44* y *VCAM1*) en rutas de interés. Evidencias tanto experimentales como predictivas relacionan también al receptor de vitamina D (*VDR)*, potencialmente asociado al peso corporal y al contenido graso (Grudberg et al. 2004) con el factor de transcripción *EGR-1,* que parece regular la ingesta de alimentos (de Lartigue et al. 2010), lo que resulta bastante congruente. Asimismo, los productos de los genes *ALDH2*, *CYP2E1*, *FASLG* y *NCK1* también aparecen relacionados con otros genes de esta red, apoyando el posible papel de estos genes en el control del crecimiento y la deposición grasa.

Los resultados obtenidos en la presente tesis doctoral abren nuevas e interesantes vías de estudio del crecimiento y la deposición grasa, aunque no tienen aplicación directa en la industria porcina. La mayor parte de los análisis realizados han permitido detectar regiones del genoma, genes y polimorfismos interesantes, muchos de los cuales no han sido estudiados hasta la fecha en la especie porcina. Por tanto, un análisis más profundo de estos así como la validación en distintas poblaciones, permitirá identificar mutaciones causales o potentes marcadores con aplicación a la selección en poblaciones comerciales. En el caso del receptor de la leptina, la asociación del polimorfismo *LEPR*c.1987C>T con caracteres de crecimiento y deposición grasa ha sido nuevamente validada, y aunque no se tienen evidencias concluyentes de su causalidad, ya que se han identificado nuevos polimorfismos que también muestran efectos sobre la expresión del mismo y no se puede descartar la existencia de más de una mutación causal, los resultados apoyan el empleo de este polimorfismo como marcador útil en programas de selección.

Perspectivas

Las tecnologías de secuenciación masiva han transformado la investigación biológica en una disciplina muy rica en datos. Cada una de las tecnologías "ómicas" (genómica, epigenómica, transcriptómica, proteómica, metabolómica, etc.) pone una pieza más para ayudarnos a completar el conocimiento de los complejos fenómenos biológicos. En el futuro próximo, dado el rápido avance de las tecnologías de análisis masivo y el abaratamiento de sus costes será posible plantear proyectos con un mayor número de muestras que los que se utilizan normalmente en la actualidad, lo que ampliará la potencia de los estudios realizados. Los continuos esfuerzos por mejorar la anotación de los genomas harán posible llevar a cabo estudios aún más completos con un mayor conocimiento de las distintas regiones del genoma. Además, la reducción en la cantidad de material genético de partida necesario para un análisis de secuenciación masiva ha permitido el desarrollo de nuevas tecnologías centradas en el análisis de una única célula, pudiendo eliminar así las etapas de amplificación por PCR, que pueden suponer un sesgo en la precisión de la determinación de los niveles de expresión.

La tecnología de los secuenciadores Helicos, PacBio y Ion Torrent permiten este tipo de secuenciación que no requiere amplificación del material genético, y son consideradas tecnologías de secuenciación de 3ª generación (McGettigan 2013).

En los últimos años han surgido, y sin duda seguirán surgiendo, dado el gran esfuerzo que está realizando la comunidad científica en este ámbito, una gran cantidad de posibilidades para explotar al máximo los resultados de los análisis de secuenciación masiva. Se han desarrollando nuevas estrategias que están permitiendo un gran avance en la detección de variantes estructurales como variaciones en el número de copias (CNVs), grandes inversiones, inserciones, delecciones y reordenamientos cromosómicos, que harán posible llevar a cabo estudios de asociación con la secuencia completa como ya han planteado otros autores (Pérez-Enciso and Ferretti 2010).

En el marco de la transcriptómica, cabe destacar las estrategias de detección de eQTLs, la expresión específica de alelos analizando los SNPs expresados (eSNPs) y el análisis de los perfiles de expresión de miRNAs y snRNAs, entre otras.

La epigenómica también está jugando un papel importante en los últimos años en el ámbito de la transcriptómica, descifrando cómo las modificaciones en DNA y las histonas controlan los perfiles de expresión génica y la regulación que comunica la transcripción y la traducción. Así han surgido los proyectos ENCODE, desarrollados por ahora en humano y ratón pero que se están empezando a aplicar en especies no modelo, con el objetivo de catalogar y diseccionar los elementos funcionales no codificantes del genoma. Los proyectos ENCODE han sido pioneros en la integración de resultados de RNA-seq, Chip-seq, DNase-seq, Ribo-seq, miRNA-seq, snoRNA-seq, etc. (Figura 13) y han llegado a la conclusión de que el 80% del genoma humano posee funciones biológicas específicas y que la mayoría de regiones no repetitivas se transcriben en ciertas circunstancias (Birney et al. 2007, Dunham et al. 2012). Además estos proyectos han permitido el desarrollo de atlas de elementos regulatorios de diferentes tejidos y líneas celulares en ratón y humano (Shen et al. 2012, Ernst et al. 2011).

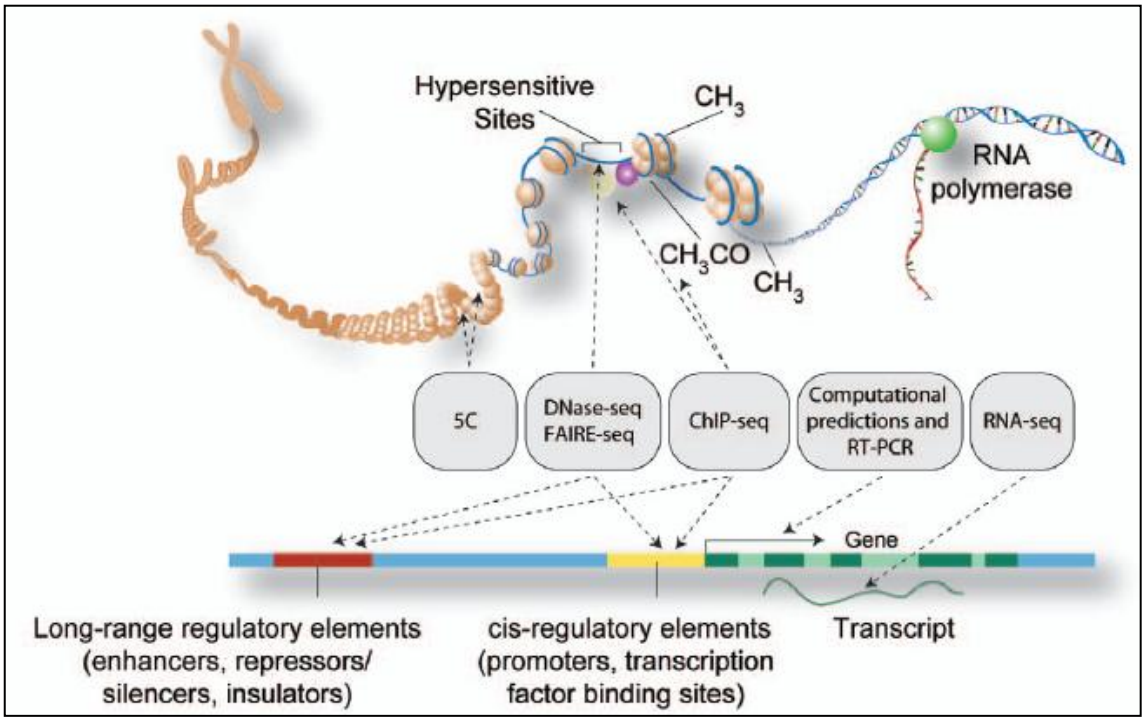


Figura 13. Representación esquemática de los principales métodos que se han usado para detectar elementos funcionales en el proyecto ENCODE (The Encode Project Consortium 2011).

La disponibilidad de esta gran cantidad de información ha permitido realizar comparaciones entre los niveles de RNA y de las proteínas resultantes, encontrándose que los niveles de tránscrito y proteína de genes individuales varían drásticamente. En general, se ha estimado en un 40% la correlación existente entre los niveles de tránscrito y proteína (Vogel and Marcotte 2012), lo que pone de manifiesto la necesidad de tener cautela al proyectar los resultados del nivel transcriptómico al dominio proteico (McGettigan 2013). Por ello es necesario, también, avanzar un paso más realizando estudios proteómicos que permitan confirmar las hipótesis planteadas a nivel genómico. Hoy en día, las tecnologías proteómicas más potentes para este tipo de estudios son la espectrometría de masas y los microarrays de proteínas. La espectrometría de masas es una tecnología muy utilizada para la identificación, cuantificación y estudios de modificaciones post-traduccionales a nivel masivo de proteínas, aunque presenta algunas desventajas en proteomas complejos y proteínas y modificaciones poco abundantes (Aebersold and Mann 2003, Desiere et al. 2005). Estas desventajas son superadas normalmente con el uso de microarrays de

proteínas que permiten llevar a cabo estudios globales del perfil proteico sin estos sesgos (Yang et al. 2011). Además los microarrays de proteínas permiten identificar interacciones proteína-proteína, proteína-ADN, proteína-ARN, proteína-lípidos, etc. generando gran cantidad de información que contribuye a comprender mejor los interactomas, es decir, estudiar de forma global cómo interactúan las proteínas en una célula para otorgar funcionalidad y sincronización a los procesos celulares (Uzoma and Zhu 2013).

El verdadero reto para los investigadores hoy en día consiste precisamente en interpretar e integrar la enorme cantidad de información generada y extraer de ella información biológica relevante. Así, han surgido en los últimos años diversas técnicas con este objetivo. La genética de sistemas es una disciplina emergente, que tiene como objetivo integrar información para explicar fenómenos biológicos desde un nivel global, descifrando el mecanismo molecular de caracteres complejos y mejorando el conocimiento de la función de los genes en las rutas bioquímicas y las interacciones genéticas (Li 2013). Es un área basada en las diferentes tecnologías "ómicas" que permite la identificación sistemática de redes génicas involucradas en el control de la expresión génica y proteíca y de interacciones entre bloques en desequilibrio de ligamiento responsables de caracteres complejos y redes metabólicas. La genética de sistemas se beneficia de la bioinformática en diversos aspectos, como el *data mining*, el conocimiento adquirido, el modelaje computacional, los análisis de rutas *in silico*, etc (Figura 14). La determinación de las redes génicas es uno de los objetivos principales de la genética de sistemas para lo que se ayuda del análisis de las correlaciones entre la expresión génica de un gran número de genes y del hecho de que genes en bloques en desequilibrio de ligamiento con patrones de expresión similares tienden a estar funcionalmente relacionados interactuando en las mismas redes (Ackermann and Beyer 2012, Weber and Hurst 2011, Bellay et al. 2011). Esta disciplina promete obtener información detallada de la relación entre las variantes génicas y los fenotipos y descifrar los mecanismos genéticos que gobiernan la independencia de múltiples fenotipos para llegar a dar respuesta a la eterna cuestión sobre la base genética de la interacción del genotipo con el ambiente (Ayroles et al. 2009).

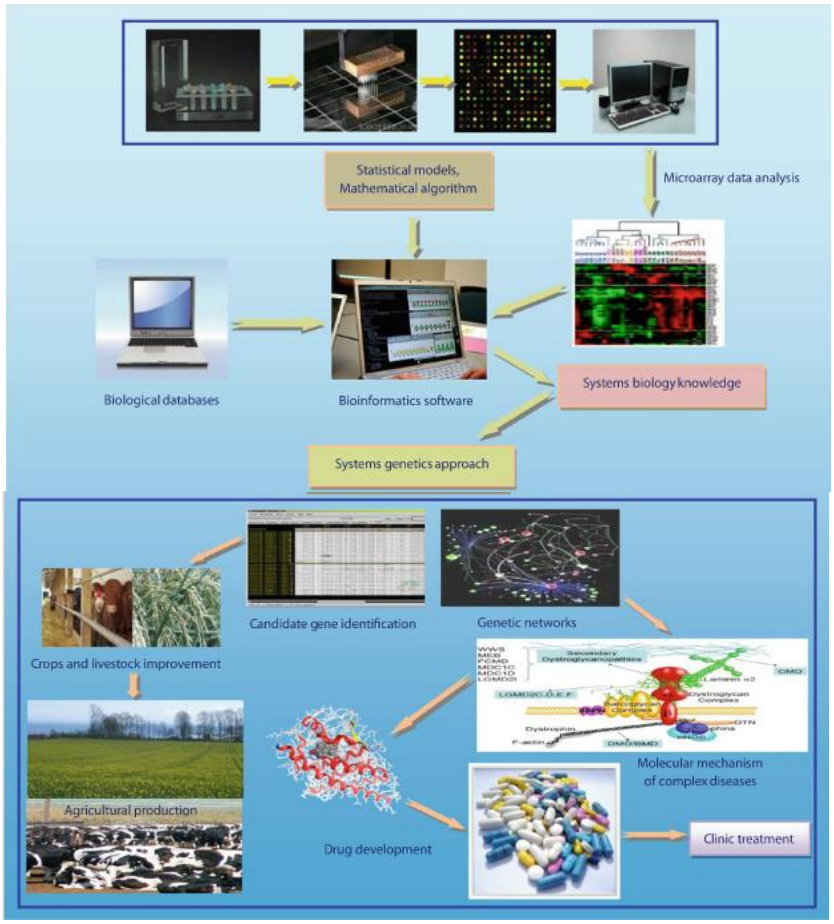


Figura 14. Genética de sistemas y su relación con la bioinformática y la biología de sistemas (Li et al. 2013).

Los avances de las tecnologías masivas de análisis genómico han dado como resultado listas casi completas de las proteínas codificadas en los genomas, que junto con los avances proteómicos y bioinformáticos han permitido la integración de todo tipo de estrategias para llegar a estudios globales de este tipo como el análisis de interactomas para conseguir llegar a traducir todo este conocimiento en un conocimiento más profundo de los sistemas biológicos y los caracteres complejos (Uzoma and Zhu 2013).

En contraposición, las estrategias clásicas basadas en el análisis de genes candidatos presentan la principal desventaja, frente a las tecnologías de análisis masivo, de requerir información a priori de estructura, función, regulación transcripcional, metabolismo, etc. que en la mayoría de los casos es limitada o inexistente (Li 2013). A pesar de todo ello, seguirán coexistiendo con las

estrategias de análisis masivo, que en algunos casos y según los objetivos, serán necesarias para la validación de los resultados y para tener un conocimiento más profundo de regiones que resulten complejas de analizar mediante otras aproximaciones.

CONCLUSIONES

1. El uso de un panel denso de marcadores ha hecho posible confirmar QTLs identificados en trabajos previos en los cromosomas SSC4 y SSC6 reduciendo sus intervalos de confianza y detectar nuevos QTLs en los cromosomas SSC1, SSC2, SSC5, SSC9, SSC11 SSC13, SSC14 y SSC17 asociados al crecimientos y la deposición grasa en porcino, pudiéndose identificar en la mayoría de ellos potentes genes candidatos posicionales y funcionales.

2. La secuenciación del transcriptoma hipotalámico mediante RNA-seq ha mostrado la elevada complejidad de este tejido comparado con otros y corroborado su importancia en la regulación del crecimiento y la deposición grasa.

3. El análisis de expresión diferencial en animales divergentes para los caracteres de interés ha permitido identificar genes, tránscritos y potenciales reguladores transcripcionales involucrados en rutas relacionadas con el control del crecimiento y la deposición grasa.

4. La secuenciación del gen *LEP* ha puesto de manifiesto la asociación del polimorfismo intrónico *LEP*g.1387C>T con caracteres de crecimiento, deposición grasa y composición corporal en un análisis conjunto con el polimorfismo *LEPR*c.1987C>T, cuyos efectos sobre estos caracteres han sido nuevamente verificados. Además se han detectado efectos interesantes de ambos polimorfismos sobre la composición de ácidos grasos en grasa subcutánea, probablemente mediados por su efecto sobre la deposición grasa.

5. El análisis estructural de la región intrónica del gen *LEP* ha evidenciado su alta variabilidad y la divergencia genética entre poblaciones asiáticas y europeas, y corrobora la localización de dos núcleos de domesticación de cerdo en Asia.

6. La nula diversidad detectada en el promotor del gen *LEPR* en la raza Ibérica podría indicar un posible evento de selección de este gen posiblemente asociado a las características particulares de esta raza en relación al apetito y deposición grasa y a las de su sistema tradicional de producción basado en el aprovechamiento de los recursos de la dehesa, de marcada estacionalidad.

7. La caracterización transcripcional del gen *LEPR* ha mostrado la existencia de una nueva isoforma corta cuya principal expresión ocurre en hígado. Los resultados parecen mostrar una regulación del *LEPR* tejido-isoforma específica.

8. Los patrones de expresión de los genes *LEP* y *LEPR* apoyan, además de su función conocida en la señalización hipotalámica, la posible implicación de ambos en funciones periféricas como el metabolismo lipídico.

CONCLUSIONS

1. The use of a dense panel of SNPs has allowed us to confirm QTLs identified in previous studies on chromosomes SSC6 and SSC4 reducing their confidence intervals and to detect new QTLs on chromosomes SSC1, SSC2, SSC5, SSC9, SSC11, SSC13, SSC14 and SSC17 associated with growth and fatness related traits in swine, identifying powerful functional and positional candidate genes in most of the regions.

2. The hypothalamic transcriptome sequencing through RNA-seq technology revealed the high complexity of this tissue in comparison to other tissues and confirmed its relevance on growth and fatness regulation.

3. The differential expression analysis performed among pigs divergent for the traits of interest has allowed us to identify genes, transcripts and potential transcriptional regulators of gene pathways involved in growth and fatness control.

4. The sequencing of the *LEP* gene has revealed an intronic polymorphism, *LEPg*.1387C>T, which is associated with growth, fatness and body composition traits in a joint analysis with *LEPRc*.1987C>T polymorphism, whose effects on these characters have been verified. In addition, interesting effects on fatty acid composition of subcutaneous fat were also detected, probably mediated through their fatness effects.

5. The structural analysis of the *LEP* gene intronic region has shown its high variability and the genetic divergence between Asian and European pig populations, and it seems to indicate the existence of two nuclei of pig domestication in Asia.

6. The null diversity detected in the promoter region of the *LEPR* gene in the Iberian breed could indicate a selection event in this gene, potentially related with the high appetite, fatness and leptin levels and to the special characteristics of Iberian traditional production system.

7. The transcriptional characterization of the *LEPR* gene has shown the existence of a new short *LEPR* isoform whose highest expression was detected in the liver. The results of this analysis seem to indicate an isoform and tissue specific regulation of this gene.

8. The expression patterns of *LEPR* and *LEP* genes support, apart from their known function in hypothalamic signaling pathways, the possible implication of both genes in peripheral functions as lipid metabolism.

REFERENCIAS

Ackermann M, Beyer A. Systematic detection of epistatic interactions based on allele pair frequencies. PLoS Genet. 2012 Feb;8(2):e1002463.

Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003 422(6928):198-207.

Ai H, Ren J, Zhang Z, Ma J, Guo Y, Yang B, Huang L. Detection of quantitative trait loci for growth- and fatness-related traits in a large-scale White Duroc × Erhualian intercross pig population. Anim Genet. 2012 43(4):383-91.

Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K, Andersson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, et al. Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science. 1994 263(5154):1771-4.

Archibald AL, Bolund L, Churcher C, Fredholm M, Groenen MA, Harlizius B, Lee KT, Milan D, Rogers J, Rothschild MF, Uenishi H, Wang J, Schook LB; Swine Genome Sequencing Consortium. Pig genome sequence--analysis and publication strategy. BMC Genomics. 2010 11:438.

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. Systems genetics of complex traits in Drosophila melanogaster. Nat Genet. 2009 41(3):299-307.

Barb C.R., G.J. Hausman, and K.L. Houseknecht. Biology of leptin in the pig. Domest. Anim. Endocrinol. 2001 21 (4): 297-317.

Bellay J, Atluri G, Sing TL, Toufighi K, Costanzo M, Ribeiro PS, Pandey G, Baller J, VanderSluis B, Michaut M, Han S, Kim P, Brown GW, Andrews BJ, Boone C, Kumar V, Myers CL. Putting genetic interactions in context through a global modular decomposition. Genome Res. 2011 21(8):1375-87.

Bidanel JP, Milan D, Iannuccelli N, Amigues Y, Boscher MY, Bourgeois F, Caritez JC, Gruand J, Le Roy P, Lagant H, Quintanilla R, Renard C, Gellin J, Ollivier L, Chevalet C. Detection of quantitative trait loci for growth and fatness in pigs. Genet Sel Evol. 2001 33(3):289-309.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D,Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R,Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S,Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S,Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P,Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED,Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D,Rosenbloom K, Kent WJ, Stone EA; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Batzoglou S, Goldman N, Hardison RC, Haussler D,Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda

MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O,Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S,Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN,Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K,Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B,McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T,Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007 447(7146):799-816.

Cabot RA, Kühholzer B, Chan AW, Lai L, Park KW, Chong KY, Schatten G, Murphy CN, Abeydeera LR, Day BN, Prather RS. Transgenic pigs produced using in vitro matured oocytes infected with a retroviral vector. Anim Biotechnol. 2001, 12(2):205-14.

Ceddia RB. Direct metabolic regulation in skeletal muscle and fat tissue by leptin: implications for glucose and fatty acids homeostasis. Int J Obes (Lond). 2005 29(10):1175-83.

Cellini E, Castellini G, Ricca V, Bagnoli S, Tedde A, Rotella CM, Faravelli C, Sorbi S, Nacmias B. Glucocorticoid receptor gene polymorphisms in Italian patients with eating disorders and obesity. Psychiatr Genet 20(6):282-8, 2010.

Chen G, Wang C, Shi T. Overview of available methods for diverse RNA-Seq data analyses. Sci China Life Sci. 2011 54(12):1121-8.

Chen CC, Chang T, Su HY. Characterization of porcine leptin receptor polymorphisms and their association with reproduction and production traits. Anim Biotechnol. 2004 15(1):89-102.

Chepelev I, Wei G, Tang Q, Zhao K. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Res. 2009 37(16):e106.

Choi I, Steibel JP, Bates RO, Raney NE, Rumph JM, Ernst CW. Application of alternative models to identify QTL for growth traits in an F2 Duroc x Pietrain pig resource population. BMC Genet. 2010;11:97.

Clop A, Ovilo C, Perez-Enciso M, Cercos A, Tomas A, Fernandez A, Coll A, Folch JM, Barragan C, Diaz I, Oliver MA, Varona L, Silio L, Sanchez A, Noguera JL. Detection of QTL affecting fatty acid composition in the pig. Mamm Genome. 2003 14(9):650-6.

Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, Pérez-Montarelo D, Noguera JL, Folch JM, Ballester M. Polymorphism in the ELOVL6 gene is associated with a major QTL effect on fatty acid composition in pigs. PLoS One. 2013; 8(1):e53687.

De Oliveira Peixoto J., S.E. Facioni, P.Savio, M.A. Menck, A. Vieira, M. V. Gualberto, R. De Almeida, E. De Almeida, and M. Silva. 2006. Associations of leptin gene polymorphisms with production traits in pigs. Journal of Animal Breeding and Genetics. 123: 378-383.

Desiere F. Towards a systems biology understanding of human health: interplay between genotype, environment and nutrition. Biotechnol Annu Rev. 2004; 10:51-84.

Duijvesteijn N, Knol EF, Merks JW, Crooijmans RP, Groenen MA, Bovenhuis H, Harlizius B. A genome-wide association study on androstenone levels in

pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet. 2010 11:42.

ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 489(7414):57-74.

ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011 9(4):e1001046.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011 473(7345):43-9.

Estellé J, Mercadé A, Pérez-Enciso M, Pena RN, Silió L, Sánchez A, Folch JM. Evaluation of FABP2 as candidate gene for a fatty acid composition QTL in porcine chromosome 8. J Anim Breed Genet. 2009 126(1):52-8.

Estellé J, Pérez-Enciso M, Mercadé A, Varona L, Alves E, Sánchez A, Folch JM. Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross. Anim Genet. 2006 37(6):589-91.

Fan B, Onteru SK, Plastow GS, Rothschild MF. Detailed characterization of the porcine MC4R gene in relation to fatness and growth. Anim Genet. 2009 40(4):401-9.

Fan B, Du ZQ, Gorbach DM and Rothschild MF. Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals. J. Anim. Sci.2010. Vol. 23, No. 7 : 833 - 847 .

Fontanesi L, Galimberti G, Calò DG, Fronza R, Martelli PL, Scotti E, Colombo M, Schiavo G, Casadio R, Buttazzoni L, Russo V. Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. J Anim Sci. 2012 90(8):2450-64.

Fontanesi L, Scotti E, Buttazzoni L, Davoli R, Russo V. The porcine fat mass and obesity associated (FTO) gene is associated with fat deposition in Italian Duroc pigs. Anim Genet. 2009 40(1):90-3.

Galve A., C. Burgos, L. Silió, L. Varona, C. Rodríguez, C. Óvilo and P. López-Buesa. 2012. The effects of leptin receptor (LEPR) and melanocortin-4 receptor (MC4R) polymorphisms on fat content, fat distribution and fat composition in a Duroc x Landrace/Large White cross. Livestock Science. 145: 145-152.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res. 2010 20(7):763-83.

Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 2009 Jn;10(6):381-91.

Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LA,Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S,Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi SH, Chow W, Clark RC, Clee C, Crooijmans RP, Dawson HD, Dehais P,De Sapio F, Dibbits B, Drou N, Du ZQ, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GJ, Fowler KE, Fredholm M, Fritz E, Gilbert JG, Giuffra E, Gorodkin J,Griffin DK, Harrow JL, Hayward A, Howe K, Hu ZL, Humphray SJ, Hunt T, Hornshøj H, Jeon JT, Jern P, Jones M, Jurka J, Kanamori H, Kapetanovic R, Kim J,Kim JH, Kim KW, Kim TH, Larson G, Lee K, Lee KT, Leggett R, Lewin HA, Li Y, Liu W, Loveland JE, Lu Y, Lunney JK, Ma J, Madsen O, Mann K, Matthews L,McLaren S, Morozumi T, Murtaugh MP, Narayan J, Nguyen DT, Ni P, Oh SJ, Onteru S, Panitz F, Park EW, Park HS, Pascal G, Paudel Y, Perez-Enciso M,Ramirez-Gonzalez R, Reecy JM, Rodriguez-Zas S, Rohrer GA, Rund L, Sang Y, Schachtschneider K, Schraiber JG, Schwartz J, Scobie L, Scott C, Searle S,Servin B, Southey BR, Sperber G, Stadler P, Sweedler JV, Tafer H, Thomsen B, Wali R, Wang J, Wang J, White S, Xu X, Yerle M, Zhang G, Zhang J, Zhang J,Zhao S, Rogers

J, Churcher C, Schook LB Analyses of pig genomes provide insight into porcine demography and evol ution. Nature 2012 491(7424):393-8.

Grundberg E, Brändström H, Ribom EL, Ljunggren O, Mallmin H, Kindmark A. Genetic variation in the human vitamin D receptor is associated with muscle strength, fat mass and body weight in Swedish women. Eur J Endocrinol. 2004 150(3):323-8.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010 28(5):503-10.

Hetzer and Harvey. Selection for High and Low Fatness in SwineJ ANIM SCI 1967 vol. 26 no. 6 1244-1251.

Hillebrand JJ, de Wied D, Adan RA. Neuropeptides, food intake and body weight regulation: a hypothalamic focus. Peptides. 2002 23(12):2283-306.

Houseknecht K.L, C.A. Baile, R.L. Matteri, and M.E. Spurlock. 1998. The biology of leptin: a review. Journal of Animal Science. 76: 1405-20.

Hu ZL, Reecy JM. Animal QTLdb: beyond a repository. A public platform for QTL comparisons and integration with diverse types of structural genomic information. Mamm Genome. 2007 18(1):1-4.

Hu J, Lou D, Carow B, Winerdal ME, Rottenberg M, Wikström AC, Norstedt G, Winqvist O. LPS regulates SOCS2 transcription in a type I interferon dependent autocrine-paracrine loop. PLoS One 2012 7(1):e30166.

Hu ZH, Park CA, Wu XL, Reecy JM. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research, 2013 41 (D1): D871-D879.

Jiang ZH, Gibson JP Genetic polymorphisms in the leptin gene and their association with fatness in four pig breeds. Mamm Genome. 1999 10(2):191-3.

Joyce AR, Palsson BØ. The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol. 2006 7(3):198-210.

Kang HS, Liao G, DeGraff LM, Gerrish K, Bortner CD, Garantziotis S, Jetten AM. CD44 plays a critical role in regulating diet-induced adipose inflammation, hepatic steatosis, and insulin resistance. PLoS One. 2013; 8(3):e58417.

Kennes YM, Murphy BD, Pothier F, Palin MF. Characterization of swine leptin (LEP) polymorphisms and their association with production traits. Anim Genet. 2001 32(4):215-8.

Kieffer TJ, Habener JF. The adipoinsular axis: effects of leptin on pancreatic beta-cells. Am J Physiol Endocrinol Metab. 2000 278(1):E1-E14.

Kim KS, Lee JJ, Shin HY, Choi BH, Lee CK, Kim JJ, Cho BW, Kim TH. Association of melanocortin 4 receptor (MC4R) and high mobility group AT-hook 1 (HMGA1) polymorphisms with pig growth and fat deposition traits. Anim Genet. 2006 37(4):419-21.

Kováčik A, Bulla J, Trakovická A, Lieskovská Z, Žitný J. Effects of the Porcine LEPR Polymorphism (Hpaii) on Carcass Traits in Large White × Landrace Crossbred Pigs. Animal Science and Biotechnologies, 2011, 44 (1).

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 2005 307(5715):1618-21.

de Lartigue G, Lur G, Dimaline R, Varro A, Raybould H, Dockray GJ. EGR1 Is a target for cooperative interactions between cholecystokinin and leptin, and inhibition by ghrelin, in vagal afferent neurons. Endocrinology 2010 151(8):3589-99.

Li H. Systems genetics in "-omics" era: current and future development. Theory Biosci. 2013 132(1):1-16.

Li X, Kim SW, Choi JS, Lee YM, Lee CK, Choi BH, Kim TH, Choi YI, Kim JJ, Kim KS. Investigation of porcine FABP3 and LEPR gene polymorphisms and mRNA

expression for variation in intramuscular fat content. Mol Biol Rep. 2010 37(8):3931-9.

Liu G, Jennen DG, Tholen E, Juengst H, Kleinwächter T, Hölker M, Tesfaye D, Un G, Schreinemachers HJ, Murani E, Ponsuksili S, Kim JJ, Schellander K,Wimmers K. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. Anim Genet. 2007 38(3):241-52.

Lu X, Liu J, Fu W, Zhou J, Luo Y, Ding X, Liu Y, Zhang Q. Genome-Wide Association Study for Cytokines and Immunoglobulin G in Swine. PLoS One. 2013 8(10):e74846.

Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C,Chinnaiyan AM. Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci U S A. 2009 106(30):12353-8.

Margetic, S., C. Gazzola, G.G. Pegg, and R.S. Hill. 2002. Leptin: a review of its peripheral actions and interactions. International Journal of Obesity. 26: 1407-1433.

Marklund L, Nyström PE, Stern S, Andersson-Eklund L, Andersson L. Confirmed quantitative trait loci for fatness and growth on pig chromosome 4. Heredity (Edinb). 1999 82 (Pt 2):134-41.

Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011 12(10):671-82. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001 157(4):1819-29.

McGettigan PA. Transcriptomics in the RNA-seq era. Curr Opin Chem Biol. 2013 17(1):4-11.

Muñoz M, Fernández AI, Benítez R, Pena RN, Folch JM, Rodríguez Mdel C, Silió L, Alves E. Disentangling two QTL on porcine chromosome 12 for backfat fatty acid composition. Anim Biotechnol. 2013; 24(3):168-86.

Muñoz M, Alves E, Corominas J, Folch JM, Casellas J, Noguera JL, Silió L, Fernández AI. Survey of SSC12 Regions Affecting Fatty Acid Composition of Intramuscular Fat Using High-Density SNP Data. Front Genet. 2012 2:101.

Muñoz G, Ovilo C, Silió L, Tomás A, Noguera JL, Rodríguez MC. Single- and joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on Sus scrofa chromosome 6 and leptin receptor effects on fatness and growth traits. J Anim Sci. 2009 87(2):459-68.

Muñoz G, Alves E, Fernández A, Ovilo C, Barragán C, Estellé J, Quintanilla R, Folch JM, Silió L, Rodríguez MC, Fernández AI. QTL detection on porcine chromosome 12 for fatty-acid composition and association analyses of the fatty acid synthase, gastric inhibitory polypeptide and acetyl-coenzyme A carboxylase alpha genes. Anim Genet. 2007 38(6):639-46.

Muñoz G, Alcázar E, Fernández A, Barragán C, Carrasco A, de Pedro E, Silió L, Sánchez JL, Rodríguez MC. Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc x Iberian crossbred pigs. 2010. Meat Sci 88:169-173.

Ollivier L. European pig genetic diversity: a minireview. Animal. 2009 3(7):915-24.

Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. Genome Biol. 2010; 11(12):220.

Óvilo C, Pérez-Enciso M, Barragan C, Clop A, Rodriguez C, Oliver MA, Toro MA, Noguera JL: A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6. Mamm Genome 2000, 11:344-346.

Óvilo C, Fernández A, Fernández AI, Folch JM, Varona L, Benítez R, Nuñez Y, Rodríguez C, Silió L. Hypothalamic expression of porcine leptin receptor (LEPR), neuropeptide Y (NPY), and cocaine- and amphetamine-regulated transcript (CART) genes is influenced by LEPR genotype. Mamm Genome. 2010 21(11-12):583-91.

Óvilo C, Fernández A, Noguera JL, Barragán C, Letón R, Rodríguez C, Mercadé A, Alves E, Folch JM, Varona L, Toro M (2005) Fine mapping of porcine

chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. Genet Res 85(1):57-67.

Óvilo C, Oliver A, Noguera JL, Clop A, Barragán C, Varona L, Rodríguez C, Toro M, Sánchez A, Pérez-Enciso M, Silió L. Test for positional candidate genes for body composition on pig chromosome 6. Genet Sel Evol. 2002 34(4):465-79.

Pérez-Enciso M, Ferretti L. Massive parallel sequencing in animal genetics: wherefroms and wheretos. Anim Genet. 2010 41(6):561-9.

Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, MacDonald TY, Cao X, Barrette T, Tewari AK, Chee MS,Chinnaiyan AM, Rickman DS, Demichelis F, Gerstein MB, Rubin MA. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome Res. 2011 21(1):56-67.

Porter V Pigs. A handbook to the breeds of the world. Mountfield (United Kingdom): Helm Information Ltd 1993.

Quintanilla R, Milan D, Bidanel JP. A further look at quantitative trait loci affecting growth and fatness in a cross between Meishan and Large White pig populations. Genet Sel Evol. 2002 34(2):193-210.

Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TP, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MA. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One. 2009 4(8):e6524.

Serra X, Gil F, Pe´rez-Enciso M, Oliver MA, Va´zquez JM et al. A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. Livest Prod Sci 1998 56, 215–223.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012 488(7409):116-20.

Shrestha YB, Wickwire K, Giraudo SQ. Action of MT-II on ghrelin-induced feeding in the paraventricular nucleus of the hypothalamus. Neuroreport. 2004 15(8):1365-7.

Siers and Thomson.Heritabilities and genetic correlations of carcass and growth traits in swine. J ANIM SCI 1972, 35:311-316.

Simpson KA, Martin NM, Bloom SR. Hypothalamic regulation of food intake and clinical therapeutic applications. Arq Bras Endocrinol Metabol. 2009 53(2):120-8.

Stanley S, Wynne K, McGowan B, Bloom S. Hormonal regulation of food intake. Physiol Rev. 2005 85(4):1131-58.

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 2008 321(5891):956-60.

Switonski M, Stachowiak M, Cieslak J, Bartz M, Grzes M. Genetics of fat tissue accumulation in pigs: a comparative approach. J Appl Genet. 2010; 51(2):153-68.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562-78, 2012.

Tyra M, Ropka-Molik K, Terman A, Piórkowska K, Oczkowicz M, Bereta A. Association between subcutaneous and intramuscular fat content in porcine ham and loin depending on age, breed and FABP3 and LEPR genes

transcript abundance. Mol Biol Rep. 2013 40(3):2301-8. doi: 10.1007/s11033-012-2311-7.

Turton MD, O'Shea D, Gunn I, Beak SA, Edwards CM, Meeran K, Choi SJ, Taylor GM, Heath MM, Lambert PD, Wilding JP, Smith DM, Ghatei MA, Herbert J,Bloom SR. A role for glucagon-like peptide-1 in the central regulation of feeding. Nature. 1996 379(6560):69-72.

Uemoto Y, Kikuchi T, Nakano H, Sato S, Shibata T, Kadowaki H, Katoh K, Kobayashi E, Suzuki K. Effects of porcine leptin receptor gene polymorphisms on backfat thickness, fat area ratios by image analysis, and serum leptin concentrations in a Duroc purebred population. Anim Sci J. 2012 83(5):375-85.

Uzoma I, Zhu H. Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. Genomics Proteomics Bioinformatics. 2013 11(1):18-28.

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature. 2003 425(6960):832-6.

Varona L, Ovilo C, Clop A, Noguera JL, Pérez-Enciso M, Coll A, Folch JM, Barragán C, Toro MA, Babot D, Sánchez A. QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects. Genet Res. 2002 80(2):145-54.

Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012 13(4):227-32.

Walling GA, Visscher PM, Andersson L, Rothschild MF, Wang L, Moser G, Groenen MA, Bidanel JP, Cepica S, Archibald AL, Geldermann H, de Koning DJ,Milan D, Haley CS. Combined analyses of data from quantitative trait loci mapping studies. Chromosome 4 effects on porcine growthand fatness. Genetics. 2000 155(3):1369-78.

Walters EM, Wolf E, Whyte JJ, Mao J, Renner S, Nagashima H, Kobayashi E, Zhao J, Wells KD, Critser JK, Riley LK, Prather RS. Completion of the swine genome will simplify the production of swine as a large animal biomedical model. BMC Med Genomics. 2012 5:55.

Wang, Y.-X., Lee, C.-H., Tiep, S., Yu, R. T., Ham, J., Kang, H., Evans, R. M. Peroxisome-proliferator-activated receptor delta activates fat metabolism to prevent obesity. Cell 2003 113: 159-170.

Weber CC, Hurst LD. Support for multiple classes of local expression clusters in Drosophila melanogaster, but no evidence for gene order conservation. Genome Biol. 2011 12(3):R23.

White S (2011) From Globalized pig breeds to capitalist pigs: a study in animals cultures and evolutionary history. Environmental History 16: 94-120.

Yadav V.K., F. Oury, K.F. Tanaka, T. Thomas, Y. Wang, S. Cremers, R: Hen, A. Krust, P. Chambon and G. Karsenty. 2011. Leptin-dependent serotonin control of appetite: temporal specificity, transcriptional regulation, and therapeutic implications. J. Exp. Med. 208 (1): 41-52.

Yang L, Guo S, Li Y, Zhou S, Tao S. Protein microarrays for systems biology. Acta Biochim Biophys Sin (Shanghai). 2011 43(3):161-71.

Zabeau L, Lavens D, Peelman F, Eyckerman S, Vandekerckhove J, Tavernier J. The ins and outs of leptin receptor activation. FEBS Lett. 2003 546(1):45-50.

ANEXO 1

Material Suplementario

ARTÍCULO I: Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip.

Supplementary Table 1 – Positions and additive effects of significant QTL at the chromosome-wide level (q-value < 0.10)

Trait	SSC	Position cM	a (SE)	*P*-value
Whole-population				
W150d	4	58	-3.38 (1.05)	1.0×10^{-3}
	7	5	3.23 (1.08)	2.8×10^{-3}
	10	27	-4.16 (1.22)	7.0×10^{-3}
BFT75	5	103	-0.47 (0.19)	1.3×10^{-2}
BFTS	4	74	0.20 (0.07)	4.1×10^{-3}
	12	91	-0.27 (0.07)	1.9×10^{-4}
	14	85	0.25 (0.07)	8.2×10^{-4}
	16	21	-0.23 (0.08)	6.2×10^{-3}
HW	1	105	-0.32 (0.10)	1.6×10^{-3}
	4	59	-0.29 (0.10)	4.4×10^{-3}
BC1 generation				
W150d	1	1	-4.51 (1.24)	3.3×10^{-4}
	2	84	3.18 (1.31)	1.5×10^{-2}
	4	58	-3.27 (1.30)	1.3×10^{-2}
	7	120	3.67 (1.25)	3.7×10^{-3}
	9	133	4.34 (1.26)	6.8×10^{-4}
	10	54	-4.49 (1.29)	5.9×10^{-4}
	13	88	3.68 (1.21)	2.5×10^{-3}
BFTS	12	91	-0.37 (0.11)	8.3×10^{-4}
	14	85	0.43 (0.11)	1.5×10^{-4}
	15	64	0.30 (0.12)	9.6×10^{-3}
	16	46	-0.28 (0.11)	4.7×10^{-3}
HW	1	96	-0.26 (0.09)	5.3×10^{-3}
	2	33	-0.22 (0.09)	2.3×10^{-2}
	13	47	0.23 (0.09)	1.4×10^{-2}
SW	4	60	-0.17 (0.05)	2.1×10^{-3}
	6	3	-0.17 (0.05)	1.5×10^{-3}
	15	14	-0.14 (0.05)	7.9×10^{-3}

	17	24	-0.15 (0.05)	4.6×10^{-3}
BLW	4	53	-0.27 (0.09)	2.1×10^{-3}
	15	91	-0.22 (0.09)	1.4×10^{-2}
	17	13	-0.22 (0.08)	9.7×10^{-3}
F3+BC2 generations				
W150d	6	111	5.3 (1.85)	4.4×10^{-3}
	9	86	-5.02 (1.75)	4.6×10^{-3}
	12	3	6.92 (1.99)	7.8×10^{-4}
ETD75	8	45	1.05 (0.33)	2.9×10^{-3}
	13	31	1.25 (0.34)	5.0×10^{-4}
	18	57	-1.43 (0.44)	2.0×10^{-3}
BFTS	11	68	0.21 (0.08)	1.1×10^{-2}
	12	50	0.28 (010)	5.0×10^{-3}
IMF	2	115	-0.45 (0.12)	3.2×10^{-4}
HW	2	115	0.99 (0.26)	5.4×10^{-4}
	4	61	-0.46 (0.18)	1.1×10^{-2}
	5	6	-0.68 (0.26)	7.8×10^{-3}
	13	18	-0.57 (0.20)	4.0×10^{-3}
BLW	4	63	-0.39 (0.12)	1.6×10^{-3}
	9	111	-0.39 (0.11)	6.2×10^{-4}

ARTÍCULO II: Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome sequencing.

Supplementary table 1. Phenotypic traits recorded from the backcrossed animals of the IBMAP population.

Description	N	Mean	SD
Average daily gain (kg/d)	162	0.728	0.11
Backfat thickness at around 90 kg (mm)	162	14.82	1.93
Percentage of C18:2 in backfat	157	14.22	1.76
Percentage of C18:2 in intramuscular fat	144	10.36	2.36

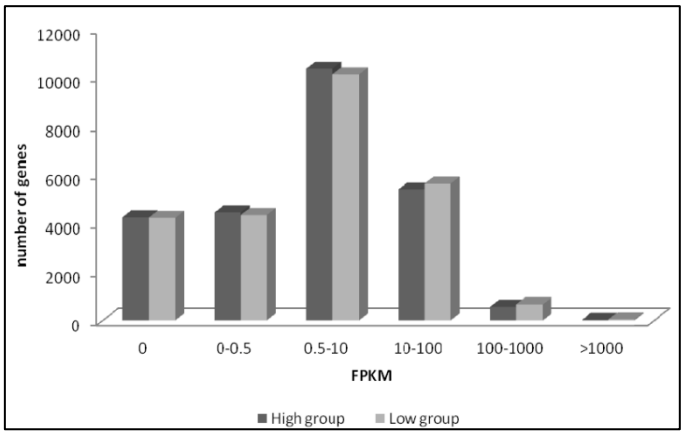
Supplementary Table 2. Total number of reads, filtered reads and percentage of mapped reads per sample. Samples marked with an asterisk are those not included in further analyses.

Sample	Total number of reads	Filtered reads	Mapped reads (%)
High_1	100,805,876	98,564,915	58.00
High_2	100,775,172	97,676,522	56.84
High_3	133,564,350	131,213,812	60.88
High_4*	103,254,478	98,162,348	60.27
High_5	98,490,020	95,578,205	55.55
Low_1*	116,030,772	112,047,709	61.94
Low_2	106,603,392	103,399,969	60.04
Low_3	105,426,048	102,701,020	62.10
Low_4	106,184,416	103,190,054	62.96
Low_5	87,573,458	84,170,362	57.68

Supplementary table 3. Primer pairs, reference sequences, size of the amplified fragment and PCR efficiency (Eff.).

transcript	Primer sequence (5' to 3')	Reference Sequence	Size bp	Eff. %
New isoform detection and quantification				
PLAC8	Fw:GACACAGCCTGCCCAGAACCTC / Rv:AGACGCCGCAGTCGCTGAA	ENSSSCT00000010120	174	92
KIAA1462	Fw:GCGCCTGCCAGCCCGAGTC / Rv:GGCACTGTCCCTTTCGCCTACAA	ENSSSCG00000011020	150	72
Reference genes quantification				
ACTB	Fw:TCTGGCACCACACCTTCT / Rv:GATCTGGGTCATCTTCTCAC	Erkens et al., 2006	114	83
B2M	Fw:TTCACACCGCTCCAGTAG / Rv:CCAGATACATAGCAGTTCAGG	Kuijk et al., 2007	166	70
TOP2B	Fw:AACTGGATGATGCTAATGATGCT / Rv:TGGAAAAACTCCGTATCTGTCTC	Erkens et al., 2006	137	74
GADPH	Fw:TCGGAGTGAACGGATTTG / Rv:CCTGGAAGATGGTGATGG	Kuijk et al., 2007	219	87
Differentially expressed genes quantification				
PRCP	Fw:GGTGCCCACTCATTCAAAGATTCC / Rv:CTGCAAGCATGCCACCGTAAGAG	ENSSSCG00000014889	163	89
ACTA2	Fw:AGAACACGGCATCATCACCAACTG / Rv:CACCGCCTGAATAGCCACATACAT	NM_001164650.1	205	81
ADAMTS4	Fw:CCGCACCCGCTTCCGTTCC / Rv:TGTAGCGAGGCACCCAGTCCAT	ENSSSCG00000006359	146	84
VAMP8	Fw:CCTGGCCCGGGGAGAAAACTTG / Rv:CCACCAGAACTTCCGAGCCACCTT	ENSSSCG00000022820	112	93
IRF1	Fw:ATCGGGCAGGACTTGGACATTGAA / Rv:TTCCCCTCCTCGTCCTCATCTGTT	NM_001097412.1	191	88
BTG2	Fw:AGGTTTTCAGCGGGGCTCTCC / Rv:CTCCCCGATGCGATAGGACACTT	NM_001097505.2	239	91
FA2H	Fw:GGGCCTCTTCGTGCTGGGGATGCT / Rv:GGGGGAAGACCAGGCGGGACTCGT	XM_003126868.3	173	84

Supplementary Figure 1. Gene expression distribution of the 25,010 genes annotated in the pig genome in FPKMs normalized values for the High (black) and Low (grey) groups.



ARTÍCULO III: Joint effects of porcine leptin and leptin receptor genes

Table S1 Phenotypic traits registered in F_2 and F_3 generations of the Iberian x Landrace cross

Description	Trait	N	Mean	SD
Growth related traits				
Weight at 150 days (kg)	W150d	358	85.05	11.38
Weight at 175 days (kg)	W175d	578	99.19	13.18
Carcass weight (kg)	CW	576	74.26	10.03
Fatness related traits				
Backfat thickness at first rib (cm)	BFT1	575	4.60	0.70
Backfat thickness at last rib (cm)	BFT2	576	2.73	0.58
Backfat thickness between third-fourth last ribs	BFT3	469	2.87	0.81
Weight of belly bacon (kg)	BBW	546	2.53	0.79
Intramuscular fat percentage (%)	IMF	508	1.51	0.58
Conformation related traits				
Mean weight of hams (kg)	HW	576	10.79	1.37
Mean weight of shoulders (kg)	SW	575	5.52	0.80
Weight of bone-in loins (kg)	BLW	576	14.86	2.28
Eye muscle area (cm^2)	EMA	453	33.97	5.02
Fatty acids composition in subcutaneous fat (%)				º
Myristic acid	C14:0	475	1.50	0.17
Butiric acid	C16:0	475	21.04	1.43
Palmitoleic acid	C16:1	475	2.49	0.39
Stearic acid	C18:0	475	10.96	1.24
Oleic acid	C18:1(n-9)	475	44.42	1.79
Vaccenic acid	C18:1(n-7)	475	3.01	0.36
Linoleic acid	C18:2	475	14.03	1.65
Linolenic acid	C18:3	475	1.08	0.19
Eicosenoic acid	C20:1	475	0.86	0.19
Eicosadienoic acid	C20:2	475	0.62	0.16
Saturated fatty acids	SFA	475	33.50	2.06
Monounsaturated fatty acids	MUFA	475	50.78	2.13
Polyunsaturated fatty acids	PUFA	475	15.73	1.80

Table S2 Primer pairs used for *LEP* sequencing and polymorphisms genotyping

Fragment	Primer sequence (5' to 3')	Size (bp)	Position (bp)	Ta (ºC)
Lep 1	Fw: tct agg ccc cag aag cac atc Rv: gaa aac acc ctg ggg aca ttg aga	633	849-1483	62
Lep 2	Fw: tga tgc cca gga cta gag gaa gc Rv: aaa ctg ggc aat ggg aag aag aac	665	1233-1896	60
Lep 3	Fw: ccc ccg ccc att gtt ct Rv: tgg agg agt cac ttg ctt tat cta	746	1802-2549	58
Lep 4	Fw: ggt gca gac agc tcc gat tag Rv: agc cac gac tgt ctg ttt ctc ac	630	2424-3059	60
Lep 5	Fw: gag act tca tga aga gcc tga cca Rv: ggt ttc ttc ccc cga ctg tgg	789	2960-1774	61
Lep 6	Fw: cca cag tcg ggg gaa gaa a Rv: gtc aca gta gaa aca aag cca gtc	611	1754-4366	61
Lep 7	Fw: agg gct ctc aag gtc gtt ctc ta Rv: cat gcc tcc ttg ttt gac cta ttg	750	4318-5069	61
Lep 8	Fw: ctt gcc gtg gcc ctc tga atg gt Rv: gat gct gat gcg ccc cga aca	661	4787-5454	63
Lep 9	Fw: gca gcc atc gca cta cag Rv: ctc cct cca ggc aat cca	781	4997-5880	59
LEPg.1382C>T *LEPg.1387C>T*	Fw: ctt cct ggt gac agc tga act ct Rv: ggg cag gac tag cag taa ggg Pyroseq: ggt gac agc tga act c	*58*	1357-1414	*63*
LEPg.1723 A>G	Fw: gga gct caa gaa tga act tgg Rv: gct gtc tgc ctt cct gta ata aa Pyroseq: aat gaa ctt ggc gaa	*75*	1697-1771	*63*

The PCR reactions were performed in a 25-µl final volume containing 150 ng of genomic DNA, standard PCR buffer [75 mM Tris-HCl pH 9.0, 50mM KCl, 20mM $(NH_4)_2$ SO_4], 2 mM of $MgCl_2$, 250 µM of deoxynucleoside triphosphate, 0.5 µM of each primer and 1 U of *Tth* DNA polymerase (Biotools, Madrid, Spain). Amplification conditions were 94ºC for 5 min., followed by 40 cycles of 94ºC (45 s), the specific annealing temperature (Table 2; 45 s) and 72ºC (45 s) with a final extension step of 10 minutes at 72ºC. The PCR reactions were carried out in an AB GeneAmp PCR System 9700 (Applied Biosystems, Warrington, UK).

Table S3 Description of the novel polymorphisms identified in *LEP* gene by sequencing the Iberian and Landrace parental pigs

Polymorphisms	Localization	Landrace alleles	Iberian alleles
g.1201C>T	Intron 1	C, T	C
g.3054A>G	Intron 1	G	A, G
g.3092A>G	Intron 1	G	A
g.3103G>A	Intron 1	A	G
g.3239A>G	Intron 1	G	A
g.3242_3244insAA	Intron 1	--	-AA-
g.4933C>T	3'UTR	T	C, T
g.5648_5650insAA	3'UTR	--	-AA-, --

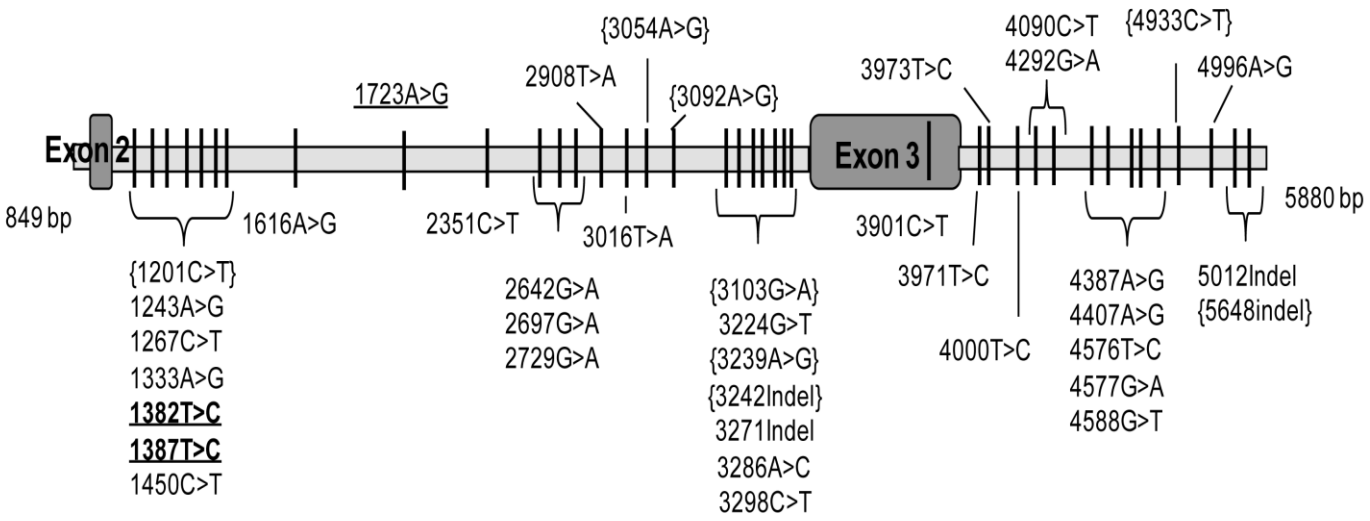
Table S4 Results of *LEP*g.1387C>T polymorphism analysis with growth, fatness and conformation traits

Trait	*LEP*g.1387C>T effects			
	Additive (SE)	*P-value*	*Dominant* (SE)	*P-value*
Growth related traits				
W150d	1.57 (0.79)	0.05	0.66 (1.09)	0.52
W175d	1.82 (0.70)	0.01	0.05 (0.90)	0.96
CW	1.22 (0.56)	0.03	-0.61 (0.74)	0.41
Fatness related traits				
BFT1	0.01 (0.04)	0.97	0.13 (0.05)	$9.0x10^{-3}$
BFT2	0.04 (0.03)	0.17	0.06 (0.04)	0.13
BFT3	0.08 (0.05)	0.10	0.18 (0.06)	$3.0x10^{-3}$
BBW	0.07 (0.03)	0.03	0.17 (0.05)	$1.8x10^{-4}$
IMF	0.04 (0.04)	0.30	0.05 (0.05)	0.30
Conformation related traits				
HW	0.04 (0.05)	0.40	0.03 (0.06)	0.65
SW	-0.01 (0.02)	0.92	-0.02 (0.03)	0.60
BLW	0.16 (0.07)	0.02	0.19 (0.09)	0.04
EMA	-0.63 (0.30)	0.03	-0.80 (0.40)	0.04
Fatty acid percentages in subcutaneous fat				
C14:0	-0.01 (0.01)	0.76	0.04 (0.02)	0.82
C16:0	-0.10 (0.08)	0.19	0.28 (0.10)	$7.0x10^{-3}$
C16:1	-0.03 (0.02)	0.24	0.08 (0.03)	0.02
C18:0	0.08 (0.08)	0.37	-0.05 (0.11)	0.68
C18:1 (n-9)	0.27 (0.11)	0.01	0.10 (0.14)	0.48
C18:1 (n-7)	-0.01 (0.02)	0.87	0.04 (0.03)	0.22
C18:2	-0.20 (0.09)	0.03	-0.43 (0.12)	$3.4 x10^{-4}$
C18:3	-0.02 (0.01)	0.16	-0.02 (0.02)	0.32
C20:1	0.01 (0.01)	0.58	-0.01 (0.02)	0.62
C20:2	-0.01 (0.01)	0.42	-0.05 (0.01)	$6.9 x10^{-4}$
SFA	-0.03 (0.12)	0.82	0.28 (0.17)	0.09
MUFA	0.25 (0.13)	0.05	0.21 (0.17)	0.22
PUFA	-0.22 (0.10)	0.02	-0.49 (0.13)	$1.8 x10^{-4}$

Table S5 Results of *LEPRc.1987C>T* polymorphism analysis with growth, fatness and conformation traits

Trait	*LEPRc.1987C>T* effects			
	Additive (SE)	*P-value*	*Dominant* (SE)	*P-value*
Growth related traits				
W150d	2.73 (0.78)	4.8×10^{-4}	-1.71 (0.99)	0.09
W175d	2.92 (0.65)	9.6×10^{-6}	-2.05 (0.86)	0.02
CW	2.18 (0.52)	3.2×10^{-5}	-0.92 (0.70)	0.15
Fatness related traits				
BFT1	0.26 (0.04)	1.1×10^{-12}	-0.05 (0.05)	0.34
BFT2	0.23 (0.03)	1.1×10^{-15}	-0.08 (0.04)	0.03
BFT3	0.37 (0.04)	1.0×10^{-16}	-0.13 (0.05)	0.02
BBW	0.24 (0.03)	2.8×10^{-14}	-0.10 (0.04)	0.02
IMF	0.20 (0.04)	9.0×10^{-8}	-0.09 (0.05)	0.05
Conformation related traits				
HW	0.04 (0.04)	0.41	-0.04 (0.06)	0.50
SW	-0.09 (0.02)	2.0×10^{-4}	0.08 (0.03)	7.3×10^{-3}
BLW	0.21 (0.07)	1.0×10^{-3}	-0.10 (0.09)	0.25
EMA	-1.31 (0.28)	5.0×10^{-6}	1.14 (0.37)	3.0×10^{-3}
Fatty acid percentages in subcutaneous fat				
C14:0	0.04 (0.01)	0.15	-0.01 (0.01)	0.87
C16:0	0.30 (0.08)	8.7×10^{-5}	-0.16 (0.10)	0.12
C16:1	0.03 (0.02)	0.23	0.03 (0.03)	0.39
C18:0	0.18 (0.08)	0.03	-0.16 (0.11)	0.13
C18:1 (n-9)	-0.19 (0.11)	0.07	0.01 (0.14)	0.92
C18:1 (n-7)	-0.02 (0.02)	0.41	0.02 (0.03)	0.46
C18:2	-0.30 (0.09)	5.1×10^{-4}	0.28 (0.11)	0.12
C18:3	-0.02 (0.01)	0.09	0.03 (0.02)	0.09
C20:1	-0.01 (0.01)	0.62	-0.05 (0.02)	4.0×10^{-3}
C20:2	-0.01 (0.01)	0.28	-0.01 (0.01)	0.71
SFA	0.52 (0.12)	1.7×10^{-5}	-0.32 (0.16)	0.04
MUFA	-0.19 (0.12)	0.13	0.02 (0.16)	0.92
PUFA	-0.33 (0.09)	4.6×10^{-4}	0.30 (0.12)	0.01

Figure S1 Scheme of the porcine *LEP* gene fragment sequenced according to the reference sequence (GenBank AJ865080.1). SNPs identified by sequencing in parental Iberian and Landrace pigs. SNPs within the same brace appeared linked in the sequenced animals. Novel SNPs appear in brackets. The SNPs selected for genotyping and association analysis appear underlined.



ARTÍCULO IV: Joint effects of porcine leptin and leptin receptor genes

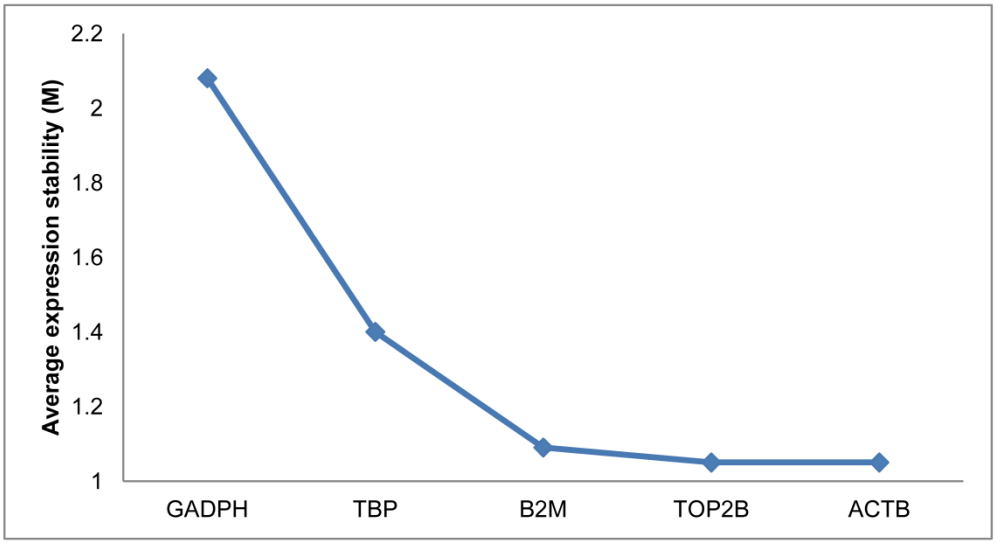
Supplementary table 1. Primer pairs, size of the amplified fragments and annealing temperature.

Fragment	Primer sequence (5' to 3')	Size, bp	Ta, ºC
LEPR promoter sequencing			
LEPRpro1	FW:TCTGCACACAGTAGGTCCTCA	514	58
	RV:CGGTCCTTTACCCAGGTGTA		
LEPRpro2	FW:CTGGGTGAAACATCCTTGGTGACA	731	58-68
	RV:GAGCGGGAAAAGGGGGAAGTAAGG		
LEPRpro3	FW:CGCCGCCGCCATCTCTG	436	58-68
	RV:GCATCCAAACCTACAAATCCTTCC		
LEP intron 2-3 sequencing			
LEPin1	FW: TGATGCCCAGGACTAGAGGAAGC	665	60
	RV: AAACTGGGCAATGGGAAGAAGAAC		
LEPin2	FW: CCCCCGCCCATTGTTCT	746	58
	RV: TGGAGGAGTCACTTGCTTTATCTA		
LEPin3	FW: GGTGCAGACAGCTCCGATTAG	630	60
	RV: AGCCACGACTGTCTGTTTCTCAC		
LEPin4	FW: GAGACT TCATGAAGAGCCTGACCA	789	61
	RV: GGTTTCTTCCCCCGACTGTGG		

Ta: annealing temperature.

ARTÍCULO V: Transcriptional characterization of porcine *leptin* and *leptin receptor* genes.

Supplementary Figure 1. Average expression stability values of the reference genes tested according to the GeNorm analysis. Those genes with stability lower than 1.5 could be considered as appropriate reference genes.



Supplementary table S1. Primer pairs and PCR conditions for expression analyses, *LEPR* isoform detection and promoters sequencing.

Fragment	Primer sequence (5' to 3')	Size, bp	Ta, ºC	Eff., %
Reference genes				
ACTB	Fw:TCTGGCACCACACCTTCT	114	60	80
	Rv:GATCTGGGTCATCTTCTCAC			
B2M	Fw:TTCACACCGCTCCAGTAG	166	60	93
	Rv:CCAGATACATAGCAGTTCAGG			
TOP2B	Fw:AACTGGATGATGCTAATGATGCT	137	60	87
	Rv:TGGAAAAACTCCGTATCTGTCTC			
GADPH	Fw:TCGGAGTGAACGGATTTG	219	60	87
	Rv:CCTGGAAGATGGTGATGG			
TBP	Fw:GATGGACGTTCGGTTTAGG	124	60	90
	Rv:AGCAGCACAGTACGAGCAA			
eEF2	Fw:CGACACTTTGAGACTGTCCAGACT Rv: AAGTGGGCCCCAGAAACC	80	60	-
Expression analyses				
LEP	Fw:GCTTTGGCCCTATCTGTCCTAC	153	60	87
	Rv:AAGTCCAAACCGGTGACCCT			
LEPRb	Fw:GAAAAACACCGGAATGATGC	239	60	81*
	Rv:AAAAGAAGAGGGCCAAATGTC			
LEPRa	Fw:AAGCTATTTTGGGAAGATGTT	139	60	96
	Rv:ATGATGGCAAGTTGGTAGATT			
LEPRglobal	Fw:GACTGGAGCACCCCCTTTACTT	207	60	98
	Rv:GCTAACATGGTCACCCACAACA			
LEPR isoform detection				
LEPR*19-20'*	Fw:GGAGTGGGGAAACCGAAGATAA	363	56	
	Rv:CATGACCAGGCAAATGACAAAG			
LEPR*5-20*	Fw:ATACAGTGCTGGATGAAAGAGGA	324	-	
	Rv:CACTGCTGCGTTGGTCACTAATA			

LEP promoter sequencing			
LEPpro1	Fw:CTCATCCGCTCCGCACCATCACC	681	62
	Rv:GGGGCAAAGCTACCGCAATCCAGA		
LEPpro2	Fw:GGTGCGGCCCCCAAAAGATAAAT	505	62
	Rv:GGCGGCTGGCAGGGATGAGAAAT		
LEPR promoter sequencing			
LEPRpro1	Fw:TCTGCACACAGTAGGTCCTCA	514	58
	Rv:CGGTCCTTTACCCAGGTGTA		
LEPRpro2	Fw:CTGGGTGAAACATCCTTGGTGACA	731	(58-68)
	Rv:GAGCGGGAAAAGGGGGAAGTAAGG		
LEPRpro3	Fw:CGCCGCCGCCATCTCTG	436	(58-68)
	Rv:GCATCCAAACCTACAAATCCTTCC		

Ta: annealing temperature; Eff: Efficiency. (*) Efficiency in hypothalamus, the only tissue where *LEPRb* could be meassured.

Supplementary table S2. Description of the polymorphisms detected in the promoter regions of *LEP* and *LEPR* genes.

Position	MAF	Predicted TF binding sites	Cosegregating group
LEPR promoter (Reference sequence FN677933.1)			
34859g.C>T	0.121	INSM1 (6.46e-01)	1
34928g.C>G	0.216	NFAT-γ-INF (9.56e-03), NFATc3 (6.03e-01) and PEA3 (7.09e-01)	2
34996g.C>T	0.121	SOX (7.08e-01)	1
35001g.T>C	0.094	SOX (7.08e-01)	3
35094g.G>T	0.216	-	2
35297g.A>G	0.243	AP-2-α/ γ (9.96e-01) and MIZ-1a (4.60e-01)	4
35373g.T>C	0.216	PEA3 (4.06e-01), MIZ-1c (4.60e-01), α -INF.2 (9.15e-01), Tst-1 consensus (7.08e-01), ε-globin (3.79e-02)	2
35384g.indel CCCCTCCACTC	0.243	MAZ (3.42e-01), NeuroD1-fp4 (3.79e-02), SP1-GPC (2.04e-02), SP1-NPY (1.43e-01), p300-consensus (1.43e-01), AP-2 (4.59e-01), SP1-cyclin-D2 (2.04e-02) and H4TF-1-FVII (2.04e-02)	4
35417g.T>C	0.243	CSRNP (4.60e-01) and Nkx-3.2 (9.75e-01)	4
35472g.A>C	0.243	-	4
35592g.G>A	0.216	γ -globin (4.60e-01)	2
35657g.G>C	0.243	AP-2 (9.15e-01), GCF (9.15e-01), SP1-complement (1.43e-01) and 57bp-URS-heptam (2.65e-01)	4
35731g.T>C	0.216	M-box (2.04e-02) and CSBP-2 (5.00e-01)	2
35782g.indel GGAGGCCCCCGGGGCGA	0.243	AP2- α / γ (9.96e-01) and MED-1 (7.09e-01)	4
35805g.A>G	0.243	LF-A1 (9.15e-01) and ELP/SF1/FTZ-F1	4

		(7.40e-02)	
35856g.G>A	0.094	CREB –IL6 (5.54e-03) and E1A (7.09e-01)	3
LEP promoter (Reference sequence AF492499)			
5112g.T>G	0.30	-	1
5127g.G>C	0.30	WT1 (6.27e-02), U-prosaposin (2.70e-02), MED-1 (5.83e-01) and AP-1-involucrin (1.03e-01)	1
5202g.C>T	0.30	PEA3 (5.83e-01), Pu box (3.54e-01) and ESE-1 (6.67e-01)	1
5344g.A>G	0.30	TBX5 (6.68e-01)	1
5374g.C>A	0.30	INSM1 (4.59e-01) and AP2- α / γ (9.80e-01)	1
5399g.T>C	0.30	-	1
5809g.T>A	0.30	EBF/Olf.1 (1.03e-01)	1

MAF: Minor allele frequency; TF: Transcription factor.

Supplementary table S3. Pairwise comparison of gene expression values observed across the five tissues tested for *LEPRglobal*, *LEPRa* and *LEP* isoforms.

Isoform	Comparison	FC	Estimator	SE	95% CI	p-value
LEPRglobal	L/BF	34.5569	-5.1109	0.3439	21.658-55.136	<.0001
	L/LD	11.9125	-3.5744	0.3444	7.461 -19.019	<.0001
	L/D	9.7297	-3.2824	0.3443	6.094 - 15.532	<.0001
	L/HT	1.9041	-0.9291	0.3680	1.154 - 3.139	0.0123
	HT/BF	18.1475	-4.1817	0.3883	10.708 - 30.755	<.0001
	HT/LD	6.2563	-2.6453	0.3888	3.689 - 10.610	<.0001
	HT/D	5.1099	-2.3533	0.3887	3.013 - 8.664	<.0001
	D/BF	3.5517	-1.8285	0.3658	2.160 - 5.838	<.0001
	D/LD	1.2243	-0.2920	0.3663	0.744 - 2.013	0.4262
	LD/BF	2.9007	-1.5364	0.3659	1.760-4.768	<.0001
LEPRa	L/BF	14.7629	-3.8839	0.3442	9.248 - 23.564	<.0001
	L/LD	6.9900	-2.8053	0.3421	4.391 - 11.125	<.0001
	L/D	5.6345	-2.4943	0.3442	3.529 - 8.993	<.0001
	L/HT	6.1509	-2.6208	0.3690	3.725 - 10.154	<.0001
	HT/BF	2.4001	-1.2631	0.3710	1.449 - 3.973	0.0008
	HT/LD	1.1364	-0.1845	0.3690	0.688 - 1.876	0.6176
	HT/D	0.9161	0.1265	0.3710	0.553 - 1.516	0.7336
	D/BF	2.6201	-1.3896	0.3463	1.636 - 4.194	<.0001
	D/LD	1.2406	-0.3110	0.3442	0.777 - 1.980	0.3673
	LD/BF	2.1120	-1.0786	0.3442	1.323 - 3.371	0.0020
LEP	BF/D	109.0390	-6.7687	0.3499	67.785 - 175.399	<.0001
	BF/LD	56.2546	-5.8139	0.3501	34.961 - 90.515	<.0001
	LD/D	1.9382	-0.9547	0.3502	1.204 - 3.119	0.0078

BF: backfat; D: diaphragm; LD: *Longissimus dorsi*; FC: fold change; SE: standard error; CI: confidence interval.

Supplementary table S4. Differential *LEP* expression conditional on *LEP*g.1387C>T genotype.

	FC	Estimator	SE	95% CI	p-value
BF CC-TT	1.2050	-0.269	0.6813	0.4775-3.0406	0.6937
BF TC-TT	0.7363	0.4416	0.6264	0.3144-1.7245	0.4823
BF CC-TC	1.6365	-0.7106	0.674	0.6550-4.0887	0.2939
BF a	1.0977	-0.1345	0.3407	0.6910-1.7438	0.6937
BF d	0.6708	0.5761	0.5543	0.3159-1.4244	0.3008
D CC-TT	0.9829	0.02494	0.6954	0.3821-2.5281	0.9715
D TC-TT	0.6643	0.5902	0.6321	0.2814-1.5678	0.3524
D CC-TC	1.4797	-0.5653	0.6649	0.5996-3.6515	0.397
D a	0.9914	0.01247	0.3477	0.6182-1.5900	0.9715
D d	0.6700	0.5777	0.5477	0.3184-1.4101	0.2936
LD CC-TT	0.8551	0.2258	0.6942	0.3330-2.1959	0.7456
LD TC-TT	0.5930	0.7539	0.6322	0.2512-1.3998	0.2355
LD CC-TC	1.4420	-0.5281	0.6663	0.5832-3.5654	0.4296
LD a	0.9247	0.1129	0.3471	0.5771-1.4819	0.7456
LD d	0.6413	0.641	0.5489	0.3042-1.3518	0.2453

BF: backfat; D: diaphragm; LD: *Longissimus dorsi*; FC: fold change: SE: standard error; CI: confidence interval.

Supplementary table S5. Differential *LEP* expression conditional on *LEP* promoter SNPs.

	FC	Estimator	SE	95% CI	Pr > |t|
BF GG-TT	1.7564	-0.8126	0.9317	0.4953-6.2278	0.3851
BF TG-TT	0.7970	0.3274	0.6158	0.3452-1.8398	0.596
BF GG-TG	2.2038	-1.1400	0.986	0.5773-8.4126	0.2502
BF a	1.3253	-0.4063	0.4659	0.7038-2.4957	0.3851
BF d	0.6014	0.7337	0.6772	0.2396-1.5090	0.2811
D GG-TT	1.7416	-0.8004	0.936	0.4883-6.2116	0.3944
D TG-TT	1.2074	-0.2719	0.6078	0.5287-2.7572	0.6555
D GG-TG	1.4423	-0.5284	0.977	0.3825-5.4389	0.5897
D a	1.3197	-0.4002	0.468	0.6988-2.4923	0.3944
D d	0.9149	0.1283	0.6655	0.3704-2.2596	0.8475
LD GG-TT	2.5448	-1.3476	0.9358	0.7137-9.0739	0.1561
LD TG-TT	0.7228	0.4682	0.6081	0.3164-1.6513	0.4458
LD GG-TG	3.5258	-1.8180	0.9774	0.9345-13.3029	0.0691
LD a	1.5952	-0.6738	0.4679	0.8448-3.0123	0.1561
LD d	0.4541	1.1390	0.6661	0.1837-1.1224	0.0927

BF: backfat; D: diaphragm; LD: *Longissimus dorsi*; FC: fold change; SE: standard error; CI: confidence interval.

ANEXO 2

Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome

OPEN ACCESS Freely available online

PLOS ONE

Porcine Tissue-Specific Regulatory Networks Derived from Meta-Analysis of the Transcriptome

Dafne Pérez-Montarelo[1,2], Nicholas J. Hudson[1], Ana I. Fernández[2], Yuliaxis Ramayo-Caldas[3], Brian P. Dalrymple[1], Antonio Reverter[1]*

1 Computational and Systems Biology, Commonwealth Scientific and Industrial Research Organisation (CSIRO) Animal, Food and Health Sciences, Queensland Bioscience Precinct, St. Lucia, Brisbane, Queensland, Australia, **2** Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain, **3** Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Abstract

The processes that drive tissue identity and differentiation remain unclear for most tissue types. So are the gene networks and transcription factors (TF) responsible for the differential structure and function of each particular tissue, and this is particularly true for non model species with incomplete genomic resources. To better understand the regulation of genes responsible for tissue identity in pigs, we have inferred regulatory networks from a meta-analysis of 20 gene expression studies spanning 480 Porcine Affymetrix chips for 134 experimental conditions on 27 distinct tissues. We developed a mixed-model normalization approach with a covariance structure that accommodated the disparity in the origin of the individual studies, and obtained the normalized expression of 12,320 genes across the 27 tissues. Using this resource, we constructed a network, based on the co-expression patterns of 1,072 TF and 1,232 tissue specific genes. The resulting network is consistent with the known biology of tissue development. Within the network, genes clustered by tissue and tissues clustered by site of embryonic origin. These clusters were significantly enriched for genes annotated in key relevant biological processes and confirm gene functions and interactions from the literature. We implemented a Regulatory Impact Factor (RIF) metric to identify the key regulators in skeletal muscle and tissues from the central nervous systems. The normalization of the meta-analysis, the inference of the gene co-expression network and the RIF metric, operated synergistically towards a successful search for tissue-specific regulators. Novel among these findings are evidence suggesting a novel key role of *ERCC3* as a muscle regulator. Together, our results recapitulate the known biology behind tissue specificity and provide new valuable insights in a less studied but valuable model species.

Citation: Pérez-Montarelo D, Hudson NJ, Fernández AI, Ramayo-Caldas Y, Dalrymple BP, et al. (2012) Porcine Tissue-Specific Regulatory Networks Derived from Meta-Analysis of the Transcriptome. PLoS ONE 7(9): e46159. doi:10.1371/journal.pone.0046159

Editor: Alberto de la Fuente, CRS4, Italy

Received May 8, 2012; **Accepted** August 28, 2012; **Published** September 26, 2012

Funding: The authors acknowledge the assistance of the CSIRO Transformational Biology Capability Platform. D. Pérez-Montarelo was funded by an FPI PhD grant from the Spanish Ministerio de Ciencia e Innovación (BES-2009-025417). Y. Ramayo-Caldas was funded by an FPU PhD grant from the Spanish Ministerio de Educación (AP2008-01450). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Tony.Reverter-Gomez@csiro.au

Introduction

Cell and tissue differentiation proceeds from tightly controlled spatial and temporal patterns of gene expression in the cell. Moreover, differences in gene expression between cell and tissue types are largely determined by transcripts derived from a limited number of tissue-specific (TS) genes, rather than by combinations of more promiscuously expressed genes [1]. Importantly, tissue specificity of gene expression has been associated with different rates of polymorphisms [2], evolution [3,4], disease association [5] and gene connectivity [6]. The identification of these TS genes is therefore likely to inform and enhance understanding of critical factors contributing to tissue specific function, structure and development. The list of transcriptional regulators driving this process is composed of transcription factors (TF), signaling molecules, co-factors, chromatin remodelers and small RNA molecules, but identifying their role in particular biological processes from expression data remains a challenge [7].

TF interact with each other to regulate the transcriptional output of a gene. However, most existing studies are focused on a limited number of TF. More often than not, it is the synergistic activity of several TF that directs the transcriptional regulation of a particular gene [8]. For this reason, the analysis of all TF interactions in a whole network appears a rational approach to better understand the complete picture of transcriptional regulation. In such a scenario, tissue-specific transcription factors (TSTF) deserve special attention, as they are the key regulators of tissue specific function and differentiation.

Here, in the spirit of meta-analysis approaches frequently invoked in genetic [9] and genomic studies [10], we integrate the data from 20 gene expression studies spanning 480 Porcine Affymetrix chips for 134 experimental conditions on 27 distinct tissues (Table 1). Analogous approaches have been undertaken before in humans, mice, cattle and other species [11,12]. Resulting from this exercise, herein we compile a matrix comprising the normalized expression of 12,320 porcine genes across 27 tissues.

We have chosen the pig, not only because of its world-wide relevance in food production, but also because it is considered as one of the most important biomedical animal models [13]. Notably, the latest instalment of the EBI Gene Expression Atlas ([14]; http://www.ebi.ac.uk/gxa) with over 19 species, does not contain the pig.

Table 1. Description of the datasets used in this study.

Reference	GEO Acc.	Chips	Tissue(s)*	Brief description
[61]	GSE26701	12	SM	4 postmortem times (20 min, 2 h, 6 h, 24 h) with 3 rep.
[62]	GSE22487	12	LD	4 developmental times (0 d, 7 d, 14 d, 21 d) with 3 rep.
[63]	GSE21383	12	OVA	6 high prolificacy replicates +6 low prolificacy rep.
[64]	GSE19975	6	LD, SOL	2 tissues with 3 rep.
[65]	GSE22165	30	BRAIN	10 conditions (3 treatments* 3/4 times) with 3 rep.
[66]	GSE18641	12	UTE	6 pregnant rep. +6 non-pregnant rep.
[67]	GSE14643	13	HEART	6 untreated rep. +7 treated rep.
[68]	GSE15256	54	ILE	3 conditions* 3 times with 6 rep.
[69]	GSE11853	12	PLA	2 breeds* 2 times with 3 rep.
[70]	GSE11787	6	SPL	3 infected rep.s +3 uninfected rep.
[71]	GSE9333	8	BFT	2 breeds with 4 rep.
[72]	GSE11193	12	LD	6 high drip loss rep. +6 low drip loss rep.
[73]	GSE7314	15	MLN	3 uninfected rep.+(3 infected rep.* 4 times)
[74]	GSE7313	15	MLN	3 uninfected rep. +(3 infected rep.* 4 times)
[75]	GSE10898	64	OLF, HYP, PIN, ADE, NEU, ACO AME, THY, DIA, BIC, BFT, AFT, STO, LIV, ILE, BLO	2 breeds* 16 tissues with 2 rep.
[76]	GSE13528	48	LIV, BFT	2 conditions* 2 genotypes* 2 tissues with 6 rep.
[77]	GSE18359	40	LIV, BFT	2 conditions* 2 RFI levels* 2 tissues with 5 rep.
[78]	GSE21096	20	HEART	4 treatments with 5 rep.
[79]	GSE23596	9	SPL	3 treatments with 3 rep.
[80]	GSE14739	80	HYP, ADE, THY, OVA, TES, BFT	4 breeds* 5 tissues with 4 rep.
TOTAL	20	480	27	

Rep.: replicates.
*Tissue codes are as follows: SM: *Semi-membranosus* muscle; LD: *Longissimus dorsi* muscle; OVA: Ovaries; SOL: *Soleus* muscle; BRAIN: Brain; UTE: Uterus; HEART: Heart; ILE: Ileum; PLA: Placenta; SPL: Spleen; BFT: Back fat tissue; MLN: Mesenteric lymph nodes; OLF: Olfactory bulb; HYP: Hypothalamus; PIN: Pineal gland; ADE: Adenohypophysis; NEU: Neurhypophysis; ACO: Adrenal cortex; AME: Adrenal medulla; LIV: Liver, THY: Thyroid gland; DIA: Diaphragm; BIC: *Biceps femoris* muscle; AFT: Abdominal fat tissue; STO: Stomach; BLO: Blood; TES: Testes.
doi:10.1371/journal.pone.0046159.t001

Further, we develop a new methodology for the identification of tissue-specific genes. This methodology analyzes the tissue of the maximum expression of each gene and maintains the distribution of maximum expressed genes observed transcriptome-wide for each particular tissue. Additionally, we present the application of the PCIT algorithm [15] to construct a tissue specific regulatory network. Finally, we describe a novel use of the regulatory impact factor (RIF) metrics [16,17] as a promising methodology for the search of TSTF in the whole transcriptome of an organism.

Results and Discussion

Quality Assessment of the Meta-Analysis Approach

The mixed-model used in the normalization accounted for 96.48% (goodness of fit, R^2) of the total variation observed in the gene expression data. Ranked from more to less relevance, the main effect of gene accounted for 59.45%, followed by the interactions of gene by tissue (23.82%), gene by experiment (9.70%) and gene by array chip (3.51%).

The normalized mean expression of 12,320 genes across 134 experimental conditions was subjected to hierarchical cluster analysis using the PermutMatrix software [18]. Multiple experimental conditions of the same tissue clustered together, confirming the validity of operating at the level of tissue after averaging across the various conditions. Table S1 provides the compiled dataset with the normalised expression of 12,320 genes across the 27 tissues. In itself, this file represents the most comprehensive atlas of the porcine transcriptome published to date. Its content was also used as the input for the PermutMatrix software to generate the hierarchical cluster analysis of tissues presented in Figure 1. The tree resulting from the hierarchical cluster analysis of 12,320 genes across the 134 conditions is given in Figure S1. The fact that tissues clustered in an anatomical and functionally sensible manner (such as the clustering of the various skeletal muscles in one branch of the hierarchical tree and tissues from the central nervous system in another branch) was attributed to the optimality of the normalization process used in the meta-analysis and anticipates the confidence in the results that emerged in the subsequent analyses.

Tissue-Specific Genes

Different methodologies for the identification of tissue-specific (TS) genes have been proposed. By and large, existing methods are direct functions of the ratio between the gene expression in one tissue to the sum total expression level across tissues. Our approach to identify TS genes combines the ratio of expressions with the distribution of the tissue location where the maximum expression of genes is observed (see Methods for details). A total of 1,234 (or 10%) of the genes were identified as TS.

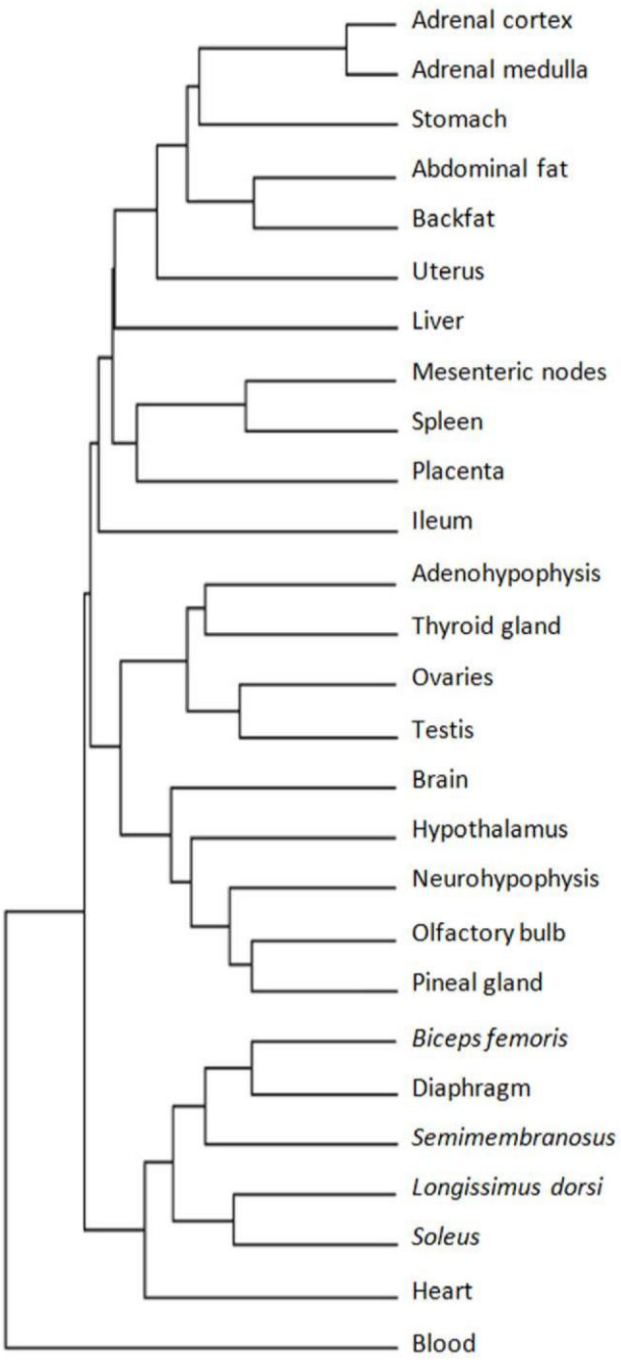


Figure 1. Clustering of tissues. Hierarchical cluster analysis of the 27 tissues based on the expression of 12,320 porcine genes.
doi:10.1371/journal.pone.0046159.g001

Figure 2A shows the distribution of the percentage of genes having its maximum expression in each tissue. By virtue of the methodology used to identify TS genes, this distribution was maintained when only the 1,232 TS genes were considered (see Materials and Methods for details in the identification of TS genes). There are noticeable differences in the proportion of genes having their maximum expression in the various tissues. On the one extreme, blood has by far the highest percentage of TS genes (14.4%) and this was attributed to blood representing a highly heterogeneous tissue with the haematopoietic cascade reported to result in the differentiation of very specific cell types [19,20]. On the other extreme, adrenal medulla is the tissue with the lowest percentage of TS genes (0.3%), followed by two muscle tissues, *Longissimus dorsi* (1.1%) and diaphragm (1.1%). It should be noted that having multiple representatives of related tissues (eg. skeletal muscle and the central nervous system (CNS), each represented by six tissues) could affect the distribution of the tissue location of the TS genes. To overcome this potential artefact, Figure 2A also shows the mean of the tissue specificity value (TSV) of the TS genes in each tissue. Although with some oscillations, it is worth noting that this value remains similar across all tissues (overlaid trend in Figure 2A), and ranges from 1.3 (for liver) to 2.2 (for abdominal fat). Importantly, the distribution of the TSV for TS genes was found to be quite different from that of the TF genes or the remaining genes (Figure 2B).While low TSV were observed for the entire set of 12,320 genes as well as for the 1,072 TF genes only, higher and more spread TSV were observed for the set of 1,230 TS genes. In this respect, 90% of all genes had a TSV ranging from 1.036 to 1.532. Similarly, 90% of all TF had a TSV ranging from 1.043 to 1.542. However, the TSV observed for 90% of TS genes ranged from 1.280 to 2.230.

A gene ontology (GO) enrichment analysis of the 1,232 TS genes (target list) against all the 12,320 genes (background list) revealed *Multicellular Organismal Process* (GO:0032501) as the most enriched biological process (*P*-value = 8.25E-17, FDR *q*-value = 8.45E-13). Moreover, the second and the third most enriched biological processes were *System Process* (GO:0003008; *P*-value = 9.11E-17, FDR *q*-value = 4.67E-13) and *Developmental Process* (GO:0032502; *P*-value = 2.52E-16, FDR *q*-value = 8.60E-13), respectively. By definition, these enriched GO terms, are related to processes whose specific outcome is the progression of cell, tissues or organs (*Multicellular Organismal Process* and *Developmental Process*) or to processes carried out by organs or tissues in multicellular organisms (*System Process*). Given that multicellular organisms are organised into tissues, this result could be a reflection of the optimality of the numerical strategy used to identify TS genes. Also, there were four muscle related GO terms in the top ten enriched processes: *Muscle Filament Sliding* (GO:0030049; *P*-value = 2.17E-15, FDR *q*-value = 4.45E-12), *Actin-Myosin Filament Sliding* (GO:0033275; I-value = 2.17E15, FDR *q*-value = 3.71E-12), *Muscle System Process* (GO:0003012; *P*-value = 6.31E-14, FDR *q*-value = 8.08E-11) and *Muscle Contraction* (GO:0006936; *P*-value = 9.86E-14, FDR *q*-value = 1.12E-10). This could be reflecting the high proportions of skeletal muscle tissue types in our data.

In the last decade, tissue specificity of gene expression has been linked to a number of important attributes including, but not limited to level of expression [21], ability to detect cis- and trans-expression quantitative trait loci [22], differential rates of polymorphism [23], imprinting [24] and evolution [3,4], as well as disease-association [5,6] and sex biased [25]. Genomic imprinting is a genetic phenomenon by which certain genes are expressed in a parent-of-origin-specific manner [26]. Table 2 shows an enrichment of transcription factors, imprinted genes and disease-associated genes among the TS genes indentified in our study. Given this prior knowledge, the results from Table 2 provide further evidence of the optimality of the analytical approach taken here to identify TS genes.

Tissue-Specific Regulatory Network

With the available 2,192 genes that included 1,120 TS, 960 TF and 112 TSTF genes, we reverse engineered a co-expression network. The overall network contained these 2,192 genes connected by 185,132 significant edges (or 7.7% of the possible 2,401,336 connections). The image of the network built from connections with significant correlation coefficients higher than 0.80 in absolute value is shown in Figure 3. This network

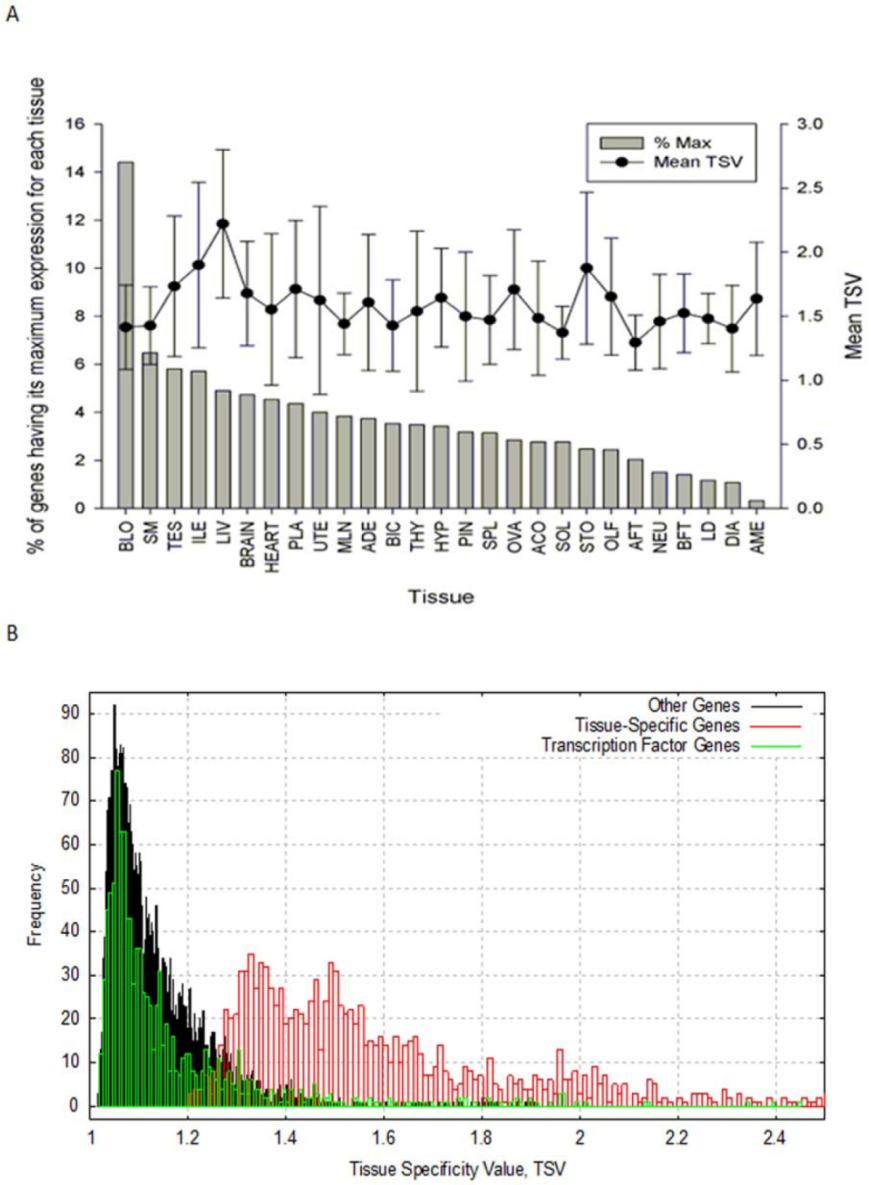


Figure 2. Tissue specificity value (TSV). (**A**) Distribution of the percentage of genes having its maximum expression in each tissue (left y axis) and the mean TSV of all the selected genes per tissue (right y axis). Standard errors are indicated as bars above and below the mean TSV. (**B**) Empirical density distribution of the TSV for tissue-specific genes (red bars), transcription factor genes (green bars) and remaining genes (black bars).
doi:10.1371/journal.pone.0046159.g002

Table 2. Enrichment of tissue-specific genes for transcription factors (TF), imprinted genes (IMP) and disease-associated genes (DIS).

	All genes (N = 12,320)		Tissue-Specific (N = 1,232)		*P*-value
	N	%[A]	N	%[A]	
TF	1,072	8.70	112	9.09	3.67E-02
IMP	134	1.09	23	1.87	3.53E-03
DIS	8,807	71.48	969	78.65	3.74E-10

[A]These percentages do not sum to one because not all of the 12,320 genes (or the subset of 1,232 tissue-specific genes) belong to one of the tree categories under scrutiny: TF, IMP and DIS.
doi:10.1371/journal.pone.0046159.t002

comprises 1,572 nodes or genes and 20,084 edges or gene connections. Figure S2 contains the Cytoscape file created to access this network.

Within the whole network, several connected components could be distinguished: one big group composed of 1,461 connected nodes, two smaller ones formed by 21 and 8 nodes respectively (on the left of the image), and a large number of small groups containing 2 to 5 genes each (showed at the bottom of the image). When the tissue where a gene had its highest expression was mapped in the visualisation schema by assigning different colours to different tissues, it became immediately apparent that nodes clustered mainly by tissue. Most of the tissues, represented by different colours in Figure 3, appeared separated from each other as independent clusters, with the exception of the six muscle tissues that clustered together in one large module (purple colours, Figure 3). An identical observation can be made for the six tissues from the central nervous system (CNS) that clustered together in their own module (green colours, Figure 3).

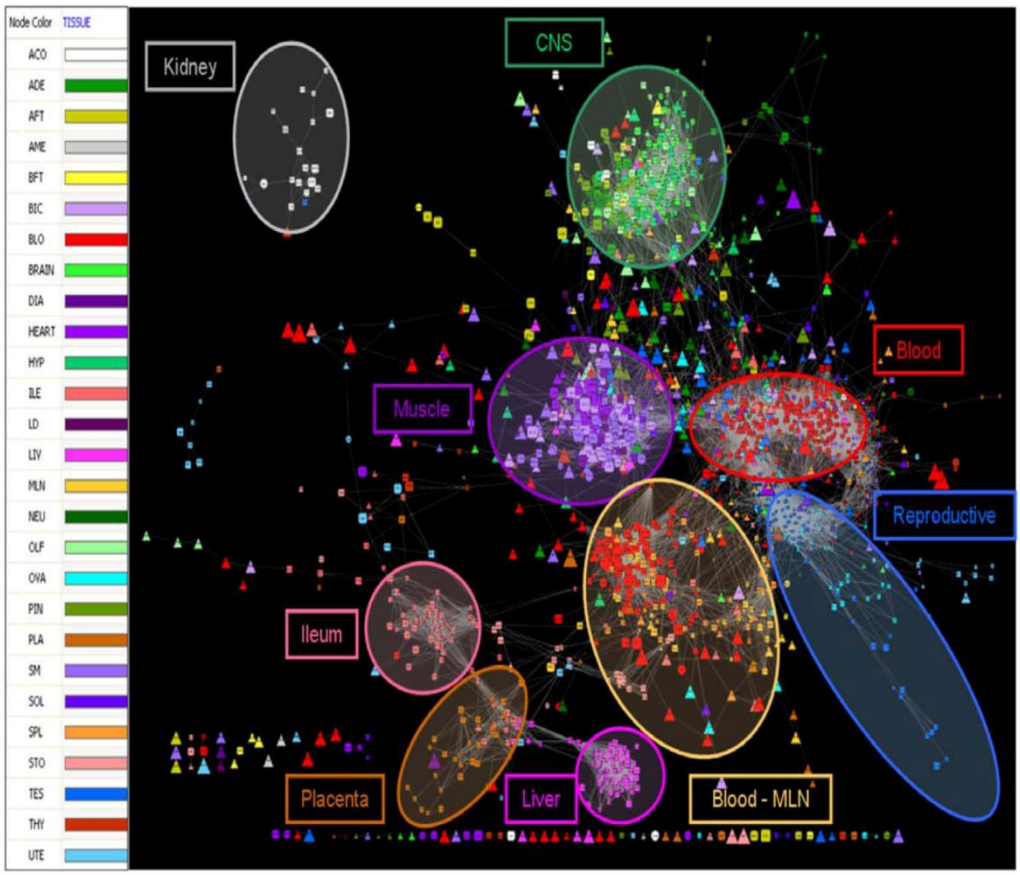


Figure 3. Tissue specific regulatory network of the porcine transcriptome. Legend of colours assigned to each of the 27 tissues in the network (left). The co-expression network (right). Node size was mapped to average transcript abundance, node colour was mapped to the tissue in which each particular gene is specific and node shape was mapped to the different gene types: TS (squares), TF (triangles) and TSTF (circles).
doi:10.1371/journal.pone.0046159.g003

The fact that each cluster represents a particular tissue was further confirmed by GO enrichment analyses. In doing so, biological processes enriched in the module assigned to '*muscle*' included *Muscle Development* (*P*-value = 1.05E-26) and *Muscle Contraction* (*P*-value = 6.17E-24). Also, the module assigned to '*CNS*' was enriched by *Nervous System Development* (*P*-value = 9.77E-12) and *Synaptic Transmission* (*P*-value = 1.53E-9). Similarly, the module formed by the mesenteric lymph nodes and the spleen was enriched for *Immune System* (*P*-value = 7.22E-17).

Moreover, if we colour each node by the embryonic origin of the tissue in which this gene is specific, it becomes apparent that tissues cluster according to their embryonic origin (Figure 4). For example, the tissues formed from the ectoderm (such as the CNS) were positioned at the top region of the network (in blue colour, Figure 4), the ones that originated from the mesoderm (case of the muscle, blood, adrenal cortex and medulla, gonads, spleen, mesenteric lymph nodes, uterus, placenta and fat) appeared in the centre of the network (in green colour, Figure 4), with the exception of the placenta that was located in the middle of the endoderm group, and the endoderm derived tissues (stomach, thyroid gland, ileum and liver) that were located in the bottom left part of the network (in yellow colour, Figure 4). Mesodermal tissues are overrepresented and more widespread in our network. Importantly, among the three germ layers, the mesoderm originated the last and its evolution is linked to the evolution of axis formation in metazoa and the appearance of eumetazoa. It is responsible for generating tissues specialized in protection, locomotion and sensing the environment that characterizes complex organisms [27].

The data and conclusions drawn from the network confirm its reliability and agreement with previous knowledge. The classification of tissues based on patterns of gene expression in the network largely reproduces classifications based on anatomical and biochemical properties [1]. Surprisingly, genes not only clustered by tissue in the network, but also, tissues clustered together according to their embryonic origin. This fact has already been noted in a mouse and human TF atlas [11] and can be attributed to these tissues being derived from transcriptional alteration of a common precursor and therefore expected to share large sections of expression patterns in common. The GO enrichment analysis provides further evidence about the quality of the inferred network and confirms that indeed it is a good representation of tissue specific regulation. Once we have confirmed the reliability of the results, we can be confident of their ability to allow for the extraction of downstream novel information about gene expression regulatory mechanisms.

Tissue-Specific Transcription Factors (TSTF)

Based on their dual classification, the 112 TSTF genes were worthy of further analyses because these type of molecules provide excellent targets for targeted tissue therapies without broadly changing other tissues. However, the reader should bear in mind that tissue specificity is a continuous variable and that not all tissues at all time points were included in this analysis. In addition, most samples contain more than one different cell type, for

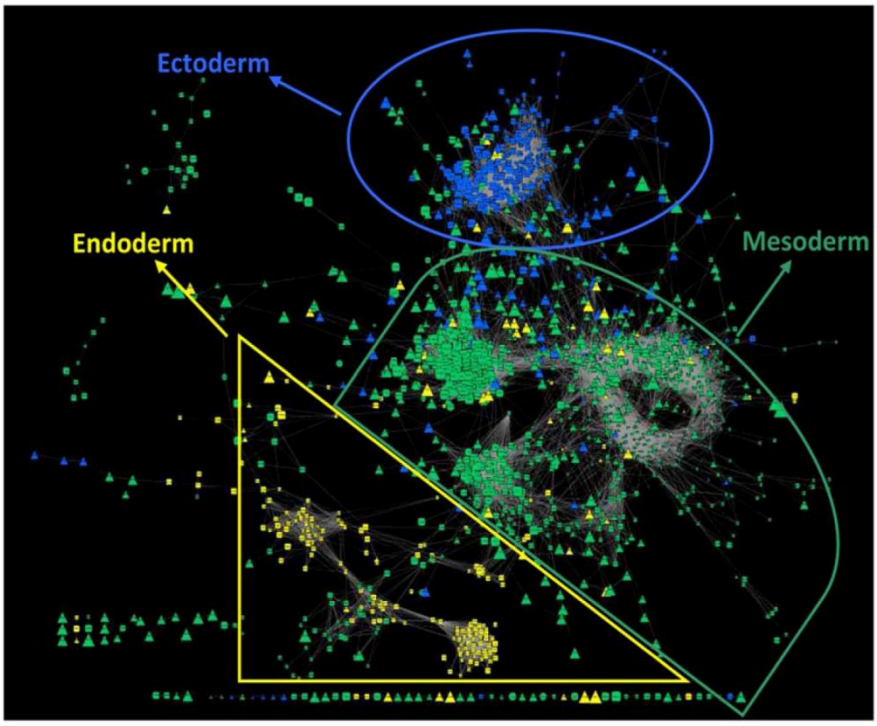


Figure 4. Embryonic origin of tissues. Tissue specific regulatory network of the porcine transcriptome, showing the embryonic origin of each tissue. In this instance, node colour was mapped to the embryonic origin of each tissue: blue for the ectoderm-derived tissues, green for the mesoderm ones and yellow for the tissues formed from the endoderm.
doi:10.1371/journal.pone.0046159.g004

example muscle contains contractile cells, adipocytes, nerve cells, blood cells and fibroblasts amongst other cell types. A search of the literature supported the tissue-specificity action for 71 of these 112 genes (Table S2). Importantly, they were distributed uniformly across the entire network, as opposed to showing a preference for a particular tissue (Figure 1 in Figure S2). When we focussed on the sub-network spanned by the TSTF genes, we identified 457 connections (Figure 5). Many interactions involving one TSTF gene were also confirmed in the literature and we highlight four. Firstly, *SIX4* appeared as a muscle specific TF in the network, directly connected to other fourteen TSTF. This TF is known to act as a regulator of *MYOD1*, the master regulator of the skeletal muscle gene expression program [28]. This role would explain the fact that *SIX4* is linked to nine muscle specific TF in the network. In addition, *SIX4* appeared also linked to four TSTF located in the CNS module of the network. Searching in the literature, we found that this could be attributed to *SIX4* playing a critical role controlling the formation of the olfactory embryonic epithelial layers and neuronal development [28]. Moreover, SIX4 carry out its action in the CNS acting synergistically with *SIX1* [29], to which *SIX4* was also connected in the network. Our second example is *KCNIP2*, a regulator of cardiac ionic currents [30] that appeared connected to several CNS genes in our network. Quite notably, *KCNIP2* plays a role in the regulation of neuronal excitability in response to intracellular ions [31]. For our third example, we emphasize *LHX9*, a pineal gland specific TF which in our network appeared connected to two testis-specific genes (*TAF7L* and *POU4F1*) and this TF has been shown to drive the axonal trajectory of some types of neurons [32] and also to play a role in gonadogenesis [33]. For our last example, we look at *GATA3*, a blood specific TF according to our tissue-specificity assignment and also related to brain and hypothalamus genes in our network. Significantly, *GATA3* has been shown to be required both in the regulation of hematopoietic stem cells [34] and in maintaining survival of the sympathetic neuron lineage [35].

All these findings support the idea that the TSTF network represents a reliable source for the generation of novel hypotheses regarding the key regulatory roles of these genes. One prominent example is the case of *GXS2*, highlighted in soft blue in Figure 5, with a total of 11 connections. This TF has not been previously associated with any reproduction or gonadal process. However, it was classified as ovary specific by our methodology, and connected in the network with the only other ovary specific gene, *BCN1*. Importantly, *BCN1* is found in abundance in the germ cells of ovaries [36]. Moreover, *GSX2* was connected to 8 more genes specific of reproductive tissues (testis and uterus). These observations support the novel hypothesis that, in addition to its known role in neuronal development in the forebrain [37], *GSX2* is a key regulator involved in gonad or reproductive processes. Similarly, *HDX*, highlighted in dark blue in Figure 5, has not been well described to date except its location on the X chromosome. In our network, this gene appeared as a testis specific TF connected with *POU4F1*, a known regulator involved in spermatogonia and expressed in distinct cell types in the testis [38], and joined to other 5 testis-specific TF. These findings suggest a potential role of *HDX* in testis development and/or function.

Muscle and CNS Transcription Regulators

In order to gain further insights into the identity of key regulators responsible for muscle and CNS differentiation and development, we undertook a series of regulatory impact factor (RIF) analyses. The aim of these analyses was to highlight those TF which, while might not be themselves differentially expressed or abundant, they still show differential connectivity, as measured by a big change in their co-expression correlation with the highly abundant highly differentially expressed genes. Figure 6 shows the

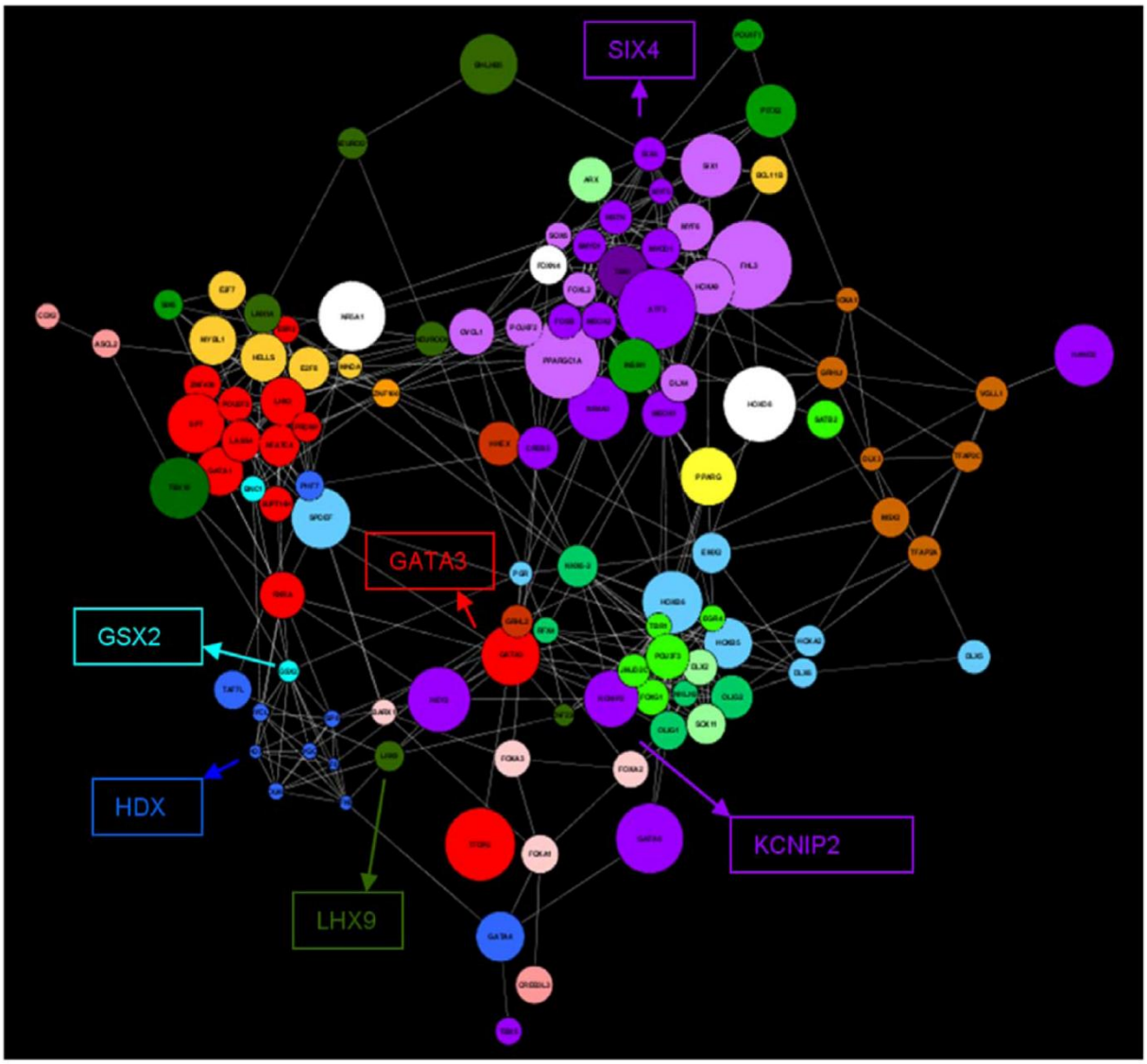


Figure 5. TSTF Network. The colour codes are as per Figure 3 and node size was mapped to average transcript abundance.
doi:10.1371/journal.pone.0046159.g005

relationship between RIF1 and RIF2 for all 1,072 TF in the two contrasts explored: 'Muscle *vs.* Other Tissues' (Figure 6A) and 'CNS *vs.* Other Tissues' (Figure 6B). The relevance of the RIF analyses became immediately clear when highlighting TF according to their tissue specificity. Muscle specific TF are highlighted in red, CNS specific TF in green and the rest of the 112 total TSTF are represented in yellow. In each contrast, the TF of biological relevance appeared preferentially located on the right half and upper-right quarter of the scatter.

Table 3 shows the results from the enrichment analysis of the TSTF of each particular tissue (muscle or CNS) when focused on the TF whose ranking is greater than 2 (based on |RIF1|+RIF2; see Methods) in each of the analyses. The ranked list of TF showed a significant enrichment of TSTF consistent with the contrast under scrutiny and more pronounced in muscle (*P*-value = 0.013) but also significant in the CNS analysis (*P*-value = 0.027). In the overall dataset, muscle-specific TF represented just a 2.5% of the total, however, when we focused on TF that showed values of |RIF1|+RIF2>2 in the 'Muscle vs. Other Tissues' comparison they represented a 5.3%. In the same way, CNS-specific TF represented only a 2.3% of the total TF, and a 4.7% of the TF with |RIF1|+RIF2>2 in the 'CNS vs. Other Tissues'' comparison. This enrichment of muscle and CNS TF in each particular analysis underscored the ability of the RIF algorithm to correctly identify the key regulators.

To further validate the performance of the RIF analyses, we searched for enriched GO terms in the ranked list of TF according to their combined RIF scores. In assessing the "Muscle *vs.* Other Tissues' output we found that 8 of the top 10 most enriched biological processes were related to muscle function or development. Some of them include: "Cell migration involved in heart development" (*P-value* = 5.25E-5, FDR *q*-value = 1.21E-1) "ventricular cardiac muscle tissue development" (*P-value* = 1.21E-4, FDR *q*-value = 1.85E-1), "muscle tissue development" (*P-value* = 7.04E-4, FDR *q*-value = 4.64E-1) and "regulation of striated muscle cell apoptotic process (*P*-value = 8.5E-4, FDR *q*-value = 3.92E-1).

The fourteen TF contained in "muscle tissue development" were ranked as follows by RIF out of the 1,072 TF (rank shown in brackets): *TBX5* (1), *SIX1* (9), *MYF6* (24), *PPP1R13L* (29), *MYOD1* (30), *GATA4* (45), *HOCXD9* (72), *MYF5* (75), *FOXP2* (116), *ZNF238* (152), *EYA2* (179), *MYOG* (196), *TCF21* (199) and *OSR1* (200). Other TF correctly prioritised by RIF, but that were overlooked by the GO enrichment analyses include *MED12* (4) [39], *MYOCD*

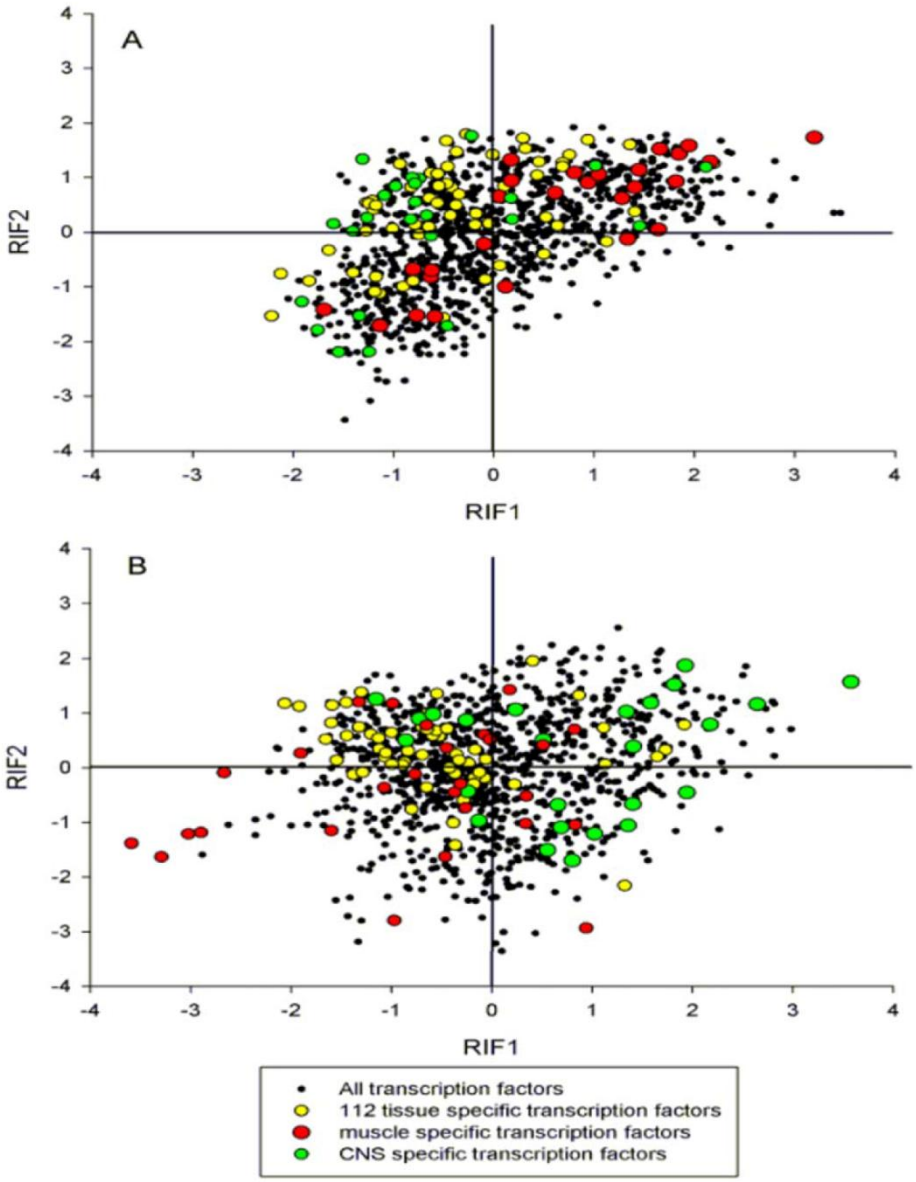


Figure 6. Regulatory impact factors (RIF). Scatter plot of the relationship between RIF1 and RIF2 in the two contrasts explored: (**A**) Muscle *vs.* Other Tissues; and (**B**) CNS *vs.* Other Tissues. Notice how in each contrast, the transcription factors of biological relevance are concentrated on the right half and upper-right quarter of the scatter.
doi:10.1371/journal.pone.0046159.g006

Table 3. Enrichment of tissue specificity in the regulatory impact factor (RIF) analysis.

	Overall		|RIF1|+RIF2>2			
			Muscle vs others		CNS vs Others	
	N	%	N	%	N	%
TF	960	89.5	151	88.3	151	88.3
TF CNS	25	2.3	3	1.7	8	4.7
TF muscle	27	2.5	9	5.3	5	3.0
TFTS	60	5.6	8	4.7	13	7.6
Total	1,072		171		177	

doi:10.1371/journal.pone.0046159.t003

(6) [40], *LMO4* (7) [41] and *PITX2* (20) [42]. The most outstanding case would be the *TBX5*, as it is one of the most extreme TF according to both RIF scores. In addition, *ERCC3* was the second gene according to RIF1 score, and it was neither assigned as a muscle-specific TF by our analyses, nor has it been previously associated to muscle function. In addition, we found that the co-expression correlation of *ERCC3* gene with *MYOCD* of −0.768 was found to be significant by the PCIT algorithm. Based on these results, we could strongly suggest a novel key role of *ERCC3* as a muscle regulator.

To gain a further insight as to the reasons why *ERCC3* scored so highly according to the RIF algorithm, we explored its relationship with the differentially expressed genes. Table 4 lists the identity of the 10 most differentially expressed genes in the 'Muscle *vs.* Other Tissues' contrast. The values for the differential co-expression of *ERCC3* and *MYOCD* with the 10 most differentially expressed genes are also given in Table 4 (differential co-expression measured by the difference in correlation co-expression in muscle

as compared to other tissues). Extreme values of differential co-expression were observed between *ERCC3* and a number of highly differentially expressed genes including *MYOZ2*, *MYOT*, *MYL1* and *TNNT1*, with differences of equivalent magnitude to those found for *MYOCD*, a well-known master regulator of cardiac and smooth muscle [40].

With respect to the RIF analysis of CNS tissues *versus* others, five out of the top ten regulators according to RIF are in fact involved in CNS development or function. The top one, *INSM1* is a neuroendocrine differentiation regulator [43], *RBL1* and *TSC22D4* are required for normal cerebellar development and differentiation [44,45], *HBP1* regulates transcription in developing myeloid cells [46] and *SIX6* acts at the hypothalamus to control reproduction and fertility [47].

We noted that the ability of the RIF analysis to identify key regulators is equally satisfactory when contemplating two very different scenarios, muscle and CNS, correctly identifying five of the top ten regulators in both cases. While the muscle cell types are all very similar between the different muscles analysed and they all have a similar function, the CNS comprises a much more complex group of tissues that includes many different cell types and functions. We could highlight, for example, the existing differences between the pineal gland, an endocrine gland responsible for secreting different hormones, and the brain, that acts as the main coordinator of the entire CNS. The performance of RIF in these two very different circumstances indicates its generality.

Concluding remarks

In conclusion, we assembled, curated and normalised a comprehensive collection of Affymetrix-based gene expression experiments in the pig in an attempt to better understand the transcriptional control of tissue development. This provided transcriptome data for 27 different tissues. Analogous approaches have been undertaken before in other species and more tissues. However, our study differs from these previous studies in two critical regards.

Firstly, we apply a set of higher-order network analyses in addition to the more conventional abundance ratio-based methods for determining tissue-specificity and tissue regulation. Because we do more than present a comprehensive survey of transcript abundances across tissues, our approach is more than a 'Gene Atlas.' Secondly, by focussing on the pig, we provide a new resource for a previously unexplored and yet important biomedical model and commercially-important livestock species.

Table 4. Normalized mean expression (NME, base-2 logarithm scale) and differential expression (DE) for the 10 most DE genes along with their co-expression correlation with *ERCC3* and *MYOCD* in the skeletal muscle and other tissues.

Gene	NME	DE	Corr. with *ERCC3* in		Corr. with *MYOCD* in	
			Muscle	Others	Muscle	Others
MYL2	6.277	4.311	−0.467	−0.596	0.449	0.591
MYOZ2	4.447	4.187	−0.579	0.402	0.638	−0.422
MYOT	4.891	4.131	0.462	0.017	−0.462	0.087
TTN	5.637	4.069	−0.118	0.163	0.237	−0.133
MB	5.086	4.061	−0.471	−0.374	0.585	0.129
CKM	7.126	3.994	−0.155	0.128	0.146	0.018
MYL1	5.396	3.989	0.823	0.162	−0.944	0.134
MYH7	7.060	3.985	−0.463	−0.409	0.445	0.099
ACTA1	7.103	3.914	−0.394	−0.611	0.555	0.651
TNNT1	5.930	3.856	0.599	0.164	−0.747	−0.328

doi:10.1371/journal.pone.0046159.t004

Our meta-analysis approach was conducted according to the preferred reporting items for systematic reviews and meta-Analysis (PRISMA; http://prisma-statement.org/). However, as it is the case with all meta-analysis approaches, our study suffers from the inability to control for experimental design effects that may contribute to bias. The data used in the present work comes from studies exploring different breeds of pigs at various developmental stages. However, we advocate that optimal normalization approaches, such as those based on mixed-model equations, allow for the integration of seemingly disparate datasets such that the results are richer in information than any of the studies taken independently.

Researchers using gene expression technologies in the quest for systems-level explanations of biological phenomena are encouraged to explore holistic measures of differential connectivity in addition to differential expression [48]. Inspired by these holistic measures, we explored a combination of strategies that allowed us to identify not only tissue-specific genes but also their transcriptional regulators. Firstly, we developed and adapted an abundance ratio metric to assess tissue-specificity. Genes highlighted by this measure are abundantly expressed in that tissue relative to others, and the approach does not skew towards any particular tissue. Then, we used both co-expression (PCIT) and differential co-expression (RIF) approaches to prioritise regulatory molecules predicted to drive each tissue phenotype (i.e., its mature physical appearance), built on the numerical foundation provided by the initial tissue-specific metric.

Through significant co-expression to tissue-specific genes, the co-expression based approaches identified important tissue regulators that may themselves be only poorly or moderately expressed in that particular tissue. On the other hand, the RIF approach identifies regulators whose behaviour (connectivity) changes between two tissues, even though they may not be strongly co-expressed in either tissue or abundantly expressed in those tissues. The PCIT co-expression network and RIF analysis exploit the same numerical signals in different ways, and therefore complement each other. We advocate the use of the combination of approaches in order to gain as much regulatory information as possible from transcriptome data.

Materials and Methods

Description of Datasets

All the datasets used in this study are publicly available microarray gene expression experiments that have been deposited on the Gene Expression Omnibus (GEO) database ([49]; http://www.ncbi.nlm.nih.gov/geo/). We have selected only those using pig tissues, that used the Affymetrix platform and that were amenable to our purpose including those surveying anatomically defined tissues. We tried to capture as many tissues as possible and without any given tissue being over-represented.

Table 1 shows the GEO accession number and a brief description of the data sets. In total, they comprise 480 Porcine Affymetrix microarray experiments from 20 independent studies. Combined, they surveyed 134 experimental conditions across 27 tissues. These 27 tissues included six muscles tissues (*Semimembranosus* (SM), *Longissimus dorsi* (LD), heart (HEART), diaphragm (DIA), *Biceps femoris* (BIC) and *Soleus* (SOL)), two fat tissues (abdominal fat tissue (AFT) and back fat tissue (BFT)), three reproductive tissues (ovaries (OVA), uterus (UTE) and testis

(TES)), two kidney regions (adrenal medulla (AME) and adrenal cortex (ACO)), two hypophysis regions (adenohypohypisis (ADE) and neurohypophysis (NEU)), two glands (thyroid gland (THY) and pineal gland (PIN)), brain (BRAIN), ileum (ILE), placenta (PLA), spleen (SPL), mesenteric lymph nodes (MLN), olfactory bulb (OLF), hypothalamus (HYP), stomach (STO), liver (LIV) and blood (BLO).

Data Processing, Quality Edits and Annotation

We obtained the MAS5 intensity signals (based-2 log expression) from all 480 microarray experiments (Table 1). The compiled data set includes many different conditions, pig breeds, ages and treatments. Therefore, a further location and scale normalization approach that took into account these features in a hierarchical fashion was deemed necessary [50]. To this effect, for each probe set (n = 24,124) the average intensity signal across biological replicates was computed resulting in 134 experimental conditions. This was followed by the computation of the average signal across the different conditions per tissue, ending with one expression value of each probe for each of the 27 tissues. The file with the normalised expression values across the 134 experimental conditions, and the file with the normalised expression values across the 27 tissues were processed by the PermutMatrix software to examine the hierarchical cluster analysis results. In those analyses we used default settings including the McQuitty's linkage method and the Euclidean distance calculation method.

Next, in order to adjust for possible heterogeneity of variance in expression signals due to tissue, we performed the Z-score normalization by tissue (ie. subtracting the tissue-specific average signal and dividing by the standard deviation of all signals in that tissue). Finally, to each z-score normalized signal we added the main effect of each probe set from its average signal across all tissues.

The original annotation of the Affymetrix Porcine chip dates from 2006 [51]. However, for the present work we used a more recent annotation (dated 2010) from the same authors and available at http://www4.ncsu.edu/~stsai2/annotation/2010-01-19_Affymetrix_Porcine_Annotation_tab_delimited.txt.

For the selection of probes, the following filtering process was applied: Firstly, we selected only those probes that were annotated to known genes. Secondly, for those genes represented by several probes sets, the probe set with the highest expression value, average across all tissues, was used as the representative of that gene, as this is supposed to be the most accurate one (i.e., high expression values tend to correspond to best quality signals). It is possible that different probe sets of the same gene represent different transcripts of that gene, but we have chosen this 'one probe set – one gene' filtering to simplify the analyses. This filtering process resulted with the z-cored normalized expression of a total of 12,320 genes across the 27 tissues (Table S1).

Further Normalization via Mixed-Model Equations

Following previously described approaches for the normalization of gene expression data with a view to co-expression analyses [50], the following linear mixed-model was fitted to the data:

$$\mathbf{y}_{ijkm} = \mathbf{A}_i + \mathbf{G}_j + \mathbf{GA}_{ij} + \mathbf{GT}_{jk} + \mathbf{GS}_{jm} + \mathbf{e}_{ijkm},$$

where $\mathbf{y}_{ijkm}$ is the vector of MAS5 z-normalized gene expression signal for the i-th array chip hybridization, from the j-th gene in the k-th tissue and m-th experimental study; $\mathbf{A}_i$ is the fixed effect of the i-th array chip hybridization ($i = 1$ to 480) and the fitting of which aims at normalizing the data by accounting for systematic non-genetic effects; $\mathbf{G}_j$ is the random component associated with the main effect of the j-th gene ($j = 1$ to 12,320); $\mathbf{GA}_{ij}$ is the random component associated with the interaction between the j-th gene and the i-th array and it captures differences from overall averages that are attributable to specific gene by array combinations; $\mathbf{GT}_{jk}$ is the random component associated with the interaction between the j-th gene and the k-th tissue ($k = 1$, to 27) and it captures differences from overall averages that are attributable to specific gene by tissue combinations; $\mathbf{GS}_{jm}$ is the random component associated with the interaction between the j-th gene and the m-th experimental study ($m = 1$ to 20) and it captures differences from overall averages that are attributable to specific gene by study combinations; and $\mathbf{e}_{ijkm}$ is the random residual error associated with $\mathbf{y}_{ijkm}$.

Using standard statistical assumptions in mixed-model theory, the effects of **G**, **GA**, **GT**, **GS** and **e** were assumed to be independent realizations from a normal distribution with zero mean and between-gene, between-gene within-array, between-gene within-tissue, between-gene within-study, and within-gene components of variance, respectively. Restricted maximum likelihood of variance components and solutions to model effects were obtained using the analytical gradients option of VCE6 software (ftp://ftp.tzv.fal.de/pub/vce6/).

Transcription Factors (TF), Imprinted and Disease-Associated Genes

Next, among the genes included in our analyses, we were interested in identifying those being TF, and/or imprinted and/or disease associated. We resorted to the census of human TF [52] to identify 1,072 TF included in our dataset. In order to identify imprinting genes included in our dataset we mined the data from the following three publicly-available gene imprinting databases: MouseBook [53] (http://www.mousebook.org/catalog.php?catalog=imprinting), Catalogue of Parent of Origin Effects [54,55] (http://igc.otago.ac.nz/home.html) and Geneimprint (http://www.geneimprint.com/site/genes-by-species). Similarly, disease-associated genes were identified as those annotated in the online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/Omim; [56]).

Identification of Tissue-Specific Genes

Different methodologies for the identification of tissue specific genes (TS) have been proposed. Most of these methods use a direct function of the ratio between the gene expression in a particular tissue to the sum total expression level across tissues [1,5]. However other tissue specificity measures that involved more complex components as the relative entropy have been presented [11]. Here, we describe a multi-tiered approach to identify TS genes. The algorithm proceeded as follows:

- Step 1: For each gene, note the tissue of its maximum expression. In formal terms, let m_i be the identity of the tissue where the i-th gene shows its maximum expression, where $i = 1, 2, \ldots, N = 12{,}320$.
- Step 2: For each tissue in j, compute p_j = the percentage of genes having its maximum expression in it. In formal terms, define

$$p_j = 100 \times \frac{\#\{m_i = ID_j\}}{N}$$

where ID_j indicates the identity of the j-th tissue and $j = 1, 2, \ldots, 27$.

- Step 3: Again for each gene, define and compute its tissue specificity value (TSV_i) as the ratio between its maximum expression and its mean expression averaged across all 27 tissues. Accordingly:

$$TSV_i = \frac{x_{m_i}}{\frac{1}{27}\sum_{j=1}^{27} x_{i_j}}$$

where x_{m_i} represents the expression of the i-th gene in tissue m_i, and x_{i_j} is the expression of the i-th gene in the j-th tissue.

- Step 4: Apply a nominal threshold of 10% by which 10% (or 1,232) of all genes in the dataset were deemed to be TS. Identify TS genes by maintaining the distribution of maximum expressions observed in the whole transcriptome. In formal terms, 10% of all p_j were assigned as TS genes after selecting based on their TSV_i.

For instance, if 6% of genes (or 740 out of 12,320) had their maximum expression in a given tissue, we selected the top 10% (or 74) of these 740 based on their TSV as TS genes. After applying this rationale to all tissues, we ended up with 10% of genes being tissue-specific and where the distribution of the location of their maximum expression was identical as that from the entire set of 12,320 genes.

Network Inference and Visualization

The expression of genes that were annotated as TF and/or TS was used to reverse-engineer a gene co-expression network using the PCIT algorithm [15]. This algorithm combines the twin concepts of partial correlation coefficient with information theory to identify significant gene to gene co-expressions, defining edges in the re-construction of the network. It works by comparing the co-expression arrangements for triplets of genes, with all triplets being exhaustively explored, and providing the correlation estimate for each pair of genes together with a flag as to whether or not the estimate has been found to be significantly different from zero. Significant correlations establish an edge in the reconstruction of the network.

Although PCIT is a soft-thresholding method, it is possible to construct networks with more or less depth using different cut offs of the absolute value of the correlation co-expression among those found to be significant. Here we present a network built with absolute co-expression correlations greater than 0.80 among those found significant by the PCIT algorithm. We have used Cytoscape version 2.6.1 [57] to visualize the network and identify modules of co-expressed genes. The organic clustering algorithm that groups together genes with common neighbours was used to visualise the topology of the network. An additional network containing only the TSTF genes was built using the orthogonal Cytoscape layout. Gene ontology (GO) enrichment analyses of modules of co-expressed genes were performed within Cytoscape using the BinGO plug-in [58].

Identification of Key Regulators: Case Study with Skeletal Muscle and the CNS

We used the Regulatory Impact Factor (RIF) metrics [16,17] to identify critical muscle and central nervous system (CNS) TF from our gene expression data. The RIF metrics identify the regulators with the highest evidence of contributing to differential expression in two biological conditions. It yields two alternate measures of TF importance, RIF1 and RIF2. The RIF1 score highlights the transcriptional regulators most differentially connected to the most abundant differentially expressed genes, while the RIF2 measure highlights those TF with the most altered ability to act as predictors of the abundance of differentially expressed genes.

While the original implementation of the RIF metrics involved the comparison of the TF with the differentially expressed genes, the exact same algebra can be adapted to the comparison of the TF with the TS genes (or any other group of genes for that matter) as long as an experimental contrast is defined (eg. Condition A *vs.* Condition B). In this respect, for the RIF analyses, we explored two contrasts: In the first one, we compared the six muscle tissues (SOL, BIC, LD, HEART, SM and DIA) against the 21 other tissues. In the second contrast, we compared the six CNS tissues (ADE, NEU, BRAIN, OLF, HYP and PIN) against the others 21 tissues. Accordingly, the RIF metrics for the i-th TF ($i = 1, 2, \ldots, 1072$) were computed using the following formulae:

$$RIF1_i = \frac{1}{n_{TS}} \sum_{j=1}^{j=n_{TS}} a_j \times d_j \times DC_{ij}^2,$$

and

$$RIF2_i = \frac{1}{n_{TS}} \sum_{j=1}^{j=n_{TS}} \left[\left(e1_j - r1_{ij}\right)^2 - \left(e2_j - r2_{ij}\right)^2 \right],$$

where n_{TS} is the number of TS genes (ie. $n_{TS} = 1{,}232$); a_j is the abundance of the j-th TS gene as given by its normalised mean expression averaged across all tissues; d_j is the differential expression of the j-th TS genes and computed from the difference between the expression of the j-th gene in the muscle minus its expression in the other tissues (for the first contrast), or from the difference between the expression of the j-th gene in the CNS minus its expression in the other tissues (for the second contrast); DC_{ij} is the differential co-expression between the i-th TF and the j-th TS gene, and computed from the difference between $r1_{ij}$, the correlation co-expression between the i-th TF and the j-th TS gene in the muscle tissues (or in the CNS tissues for the second contrast), and $r2_{ij}$, the correlation co-expression between the i-th TF and the j-th TS gene in the remaining tissues; and $e1_j$ and $e2_j$ represent the normalised mean expression of the j-th TS gene averaged across all muscle tissues (or across all CNS tissues for the second contrast) and across all the remaining tissues, respectively.

Importantly, RIF1 depends on the direction of (or which condition is used first in) the contrast, "A *versus* B" or "B *versus* A". Instead, the sign of RIF2 is not affected by this contrast directionality, but by the change in the ability of the TF to predict the abundance of DE in the two conditions, regardless of which condition is considered first in the contrast. For this reason, we ranked TF based on their |RIF1|+RIF2 score. Finally, the ranked list of TF was processed through the GOrilla tool [59] to search for enriched GO terms. From this tool, we report the enrichment *P*-value computed from the hypergeometric test and the false discovery rate (FDR) *q*-value which corresponds to the p-value corrected for multiple testing using the Benjamini and Hochberg method [60].

Supporting Information

Figure S1 Hierarchical cluster analysis of the 143 experimental conditions based on the expression of the 12,320 porcine genes.
(TIF)

Figure S2 Cytoscape formatted file to allow the visualization and recreation of the networks presented in this study.
(GZ)

Table S1 Comma delimited file with the normalized mean expression of 12,320 genes across the 27 tissues.
(CSV)

Table S2 Word document file listing the 112 tissue specific transcription factor genes, their expression and their location in the network.
(DOC)

References

1. Jongeneel CV, Delorenzi M, Iseli C, Zhou DX, Haudenschild CD, et al. (2005) An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Research 15: 1007–1014.
2. Zhang LQ, Li WH (2005) Human SNPs reveal no evidence of frequent positive selection. Molecular Biology and Evolution 22: 2504–2507.
3. Nagaraj SH, Ingham A, Reverter A (2010) The interplay between evolution, regulation and tissue specificity in the Human Hereditary Diseasome. BMC Genomics 11 Suppl 4: S23.
4. Yang J, Su AI, Li WH (2005) Gene expression evolves faster in narrowly than in broadly expressed mammalian genes. Molecular Biology and Evolution 22: 2113–2118.
5. Winter EE, Goodstadt L, Ponting CP (2004) Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. Genome Research 14: 54–61.
6. Reverter A, Ingham A, Dalrymple BP (2008) Mining tissue specificity, gene connectivity and disease association to reveal a set of genes that modify the action of disease causing genes. BioData mining 1: 8.
7. Fedorova E, Zink D (2008) Nuclear architecture and gene regulation. Biochimica Et Biophysica Acta-Molecular Cell Research 1783: 2174–2184.
8. Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, et al. (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. Mol Syst Biol 5: 294.
9. Zeggini E, Ioannidis JPA (2009) Meta-analysis in genome-wide association studies. Pharmacogenomics 10: 191–201.
10. Cahan P, Rovegno F, Mooney D, Newman JC, St. Laurent G, III, et al. (2007) Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. Gene 401: 12–18.
11. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, et al. (2010) An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man (vol 140, pg 744, 2010). Cell 141: 369–369.
12. Harhay GP, Smith TP, Alexander LJ, Haudenschild CD, Keele JW, et al. (2010) An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. Genome Biol 11: R102.
13. Swindle MM, Makin A, Herron AJ, Clubb FJ Jr, Frazier KS (2011) Swine as Models in Biomedical Research and Toxicology Testing. Vet Pathol.
14. Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farne A, et al. (2011) Gene Expression Atlas update–a value-added database of microarray and sequencing-based functional genomics experiments. Nucleic Acids Res.
15. Reverter A, Chan EKF (2008) Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics 24: 2491–2497.
16. Hudson NJ, Reverter A, Dalrymple BP (2009) A Differential Wiring Analysis of Expression Data Correctly Identifies the Gene Containing the Causal Mutation. Plos Computational Biology 5.
17. Reverter A, Hudson NJ, Nagaraj SH, Perez-Enciso M, Dalrymple BP (2010) Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics 26: 896–904.
18. Caraux G, Pinloche S (2005) PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics 21: 1280–1281.
19. Cedar H, Bergman Y (2011) Epigenetics of haematopoietic cell development. Nat Rev Immunol 11: 478–488.
20. Pimanda JE, Gottgens B (2010) Gene regulatory networks governing haematopoietic stem cell development and identity. Int J Dev Biol 54: 1201–1211.
21. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101: 6062–6067.
22. Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, et al. (2006) Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet 2: e172.
23. Zhang L, Li WH (2005) Human SNPs reveal no evidence of frequent positive selection. Mol Biol Evol 22: 2504–2507.
24. Ideraabdullah FY, Vigneau S, Bartolomei MS (2008) Genomic imprinting mechanisms in mammals. Mutat Res 647: 77–85.
25. Meisel RP, Malone JH, Clark AG (2012) Disentangling the relationship between sex-biased gene expression and X-linkage. Genome Res.

Acknowledgments

The authors are grateful to Dr. Gene Wiffels and Dr. Aaron Ingham for their insightful review of this manuscript. We thank Dr. Cedric Gondro for assistance unpacking some of the gene expression datasets used in the study.

Author Contributions

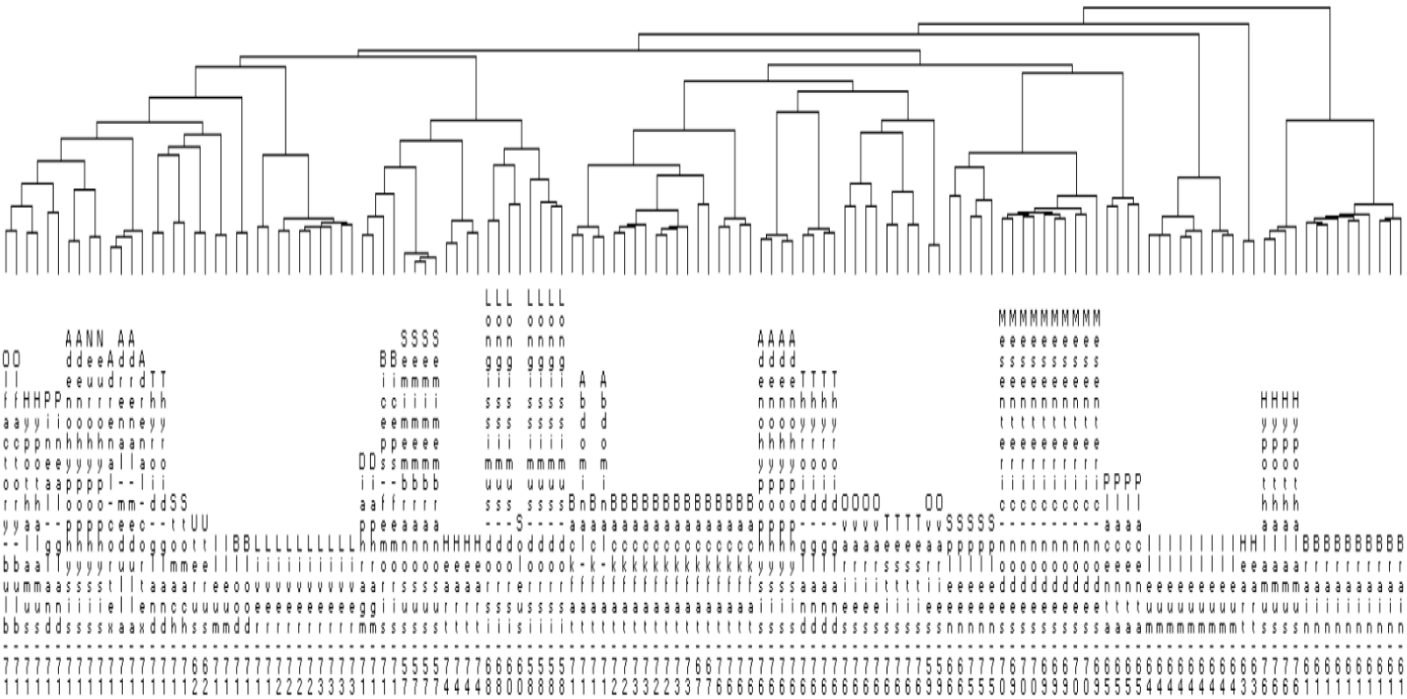
Conceived and designed the experiments: DPM NJH AR. Performed the experiments: DPM AR. Analyzed the data: DPM NJH AIF YRC BPD AR. Contributed reagents/materials/analysis tools: DPM NJH AIF YRC BPD AR. Wrote the paper: DPM NJH AIF YRC BPD AR.

26. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2: 21–32.
27. Technau U, Scholz CB (2003) Origin and evolution of endoderm and mesoderm. Int J Dev Biol 47: 531–539.
28. Aziz A, Liu Q-C, Dilworth FJ (2010) Regulating a master regulator Establishing tissue-specific gene expression in skeletal muscle. Epigenetics 5: 691–695.
29. Chen B, Kim E-H, Xu P-X (2009) Initiation of olfactory placode development and neurogenesis is blocked in mice lacking both Six1 and Six4. Developmental Biology 326: 75–85.
30. Thomsen MB, Wang C, Ozgen N, Wang H-G, Rosen MR, et al. (2009) Accessory Subunit KChIP2 Modulates the Cardiac L-Type Calcium Current. Circulation Research 104: 1382–U1251.
31. Norris AJ, Foeger NC, Nerbonne JM (2010) Interdependent Roles for Accessory KChIP2, KChIP3, and KChIP4 Subunits in the Generation of Kv4-Encoded I(A) Channels in Cortical Pyramidal Neurons. Journal of Neuroscience 30: 13644–13655.
32. Avraham O, Hadas Y, Vald L, Zisman S, Schejter A, et al. (2009) Transcriptional control of axonal guidance and sorting in dorsal interneurons by the Lim-HD proteins Lhx9 and Lhx1. Neural Development 4.
33. Mazaud S, Oreal E, Guigon CJ, Carre-Eusebe D, Magre S (2002) Lhx9 expression during gonadal morphogenesis as related to the state of cell differentiation. Gene expression patterns : GEP 2: 373–377.
34. Buza-Vidas N, Duarte S, Luc S, Bouriez-Jones T, Woll PS, et al. GATA3 is redundant for maintenance and self-renewal of hematopoietic stem cells. Blood 118: 1291–1293.
35. Tsarovina K, Reiff T, Stubbusch J, Kurek D, Grosveld FG, et al. (2010) The Gata3 Transcription Factor Is Required for the Survival of Embryonic and Adult Sympathetic Neurons. Journal of Neuroscience 30: 10833–10843.
36. Mahoney MG, Tang W, Xiang MM, Moss SB, Gerton GL, et al. (1998) Translocation of the zinc finger protein basonuclin from the mouse germ cell nucleus to the midpiece of the spermatozoon during spermiogenesis. Biology of Reproduction 59: 388–394.
37. Pei Z, Wang B, Chen G, Nagao M, Nakafuku M, et al. (2011) Homeobox genes Gsx1 and Gsx2 differentially regulate telencephalic progenitor maturation. Proc Natl Acad Sci U S A 108: 1675–1680.
38. Budhram-Mahadeo V, Moore A, Morris PJ, Ward T, Weber B, et al. (2001) The closely related POU family transcription factors Brn-3a and Brn-3b are expressed in distinct cell types in the testis. International Journal of Biochemistry & Cell Biology 33: 1027–1039.
39. Rocha PP, Scholze M, Bleiss W, Schrewe H (2010) Med12 is essential for early mouse development and for canonical Wnt and Wnt/PCP signaling. Development 137: 2723–2731.
40. Ilagan RM, Genheimer CW, Quinlan SF, Guthrie KI, Sangha N, et al. (2011) Smooth Muscle Phenotypic Diversity Is Mediated Through Alterations in Myocardin Gene Splicing. Journal of Cellular Physiology 226: 2702–2711.
41. Chen HH, Yip JW, Stewart AF, Frank E (2002) Differential expression of a transcription regulatory factor, the LIM domain only 4 protein Lmo4, in muscle sensory neurons. Development 129: 4879–4889.
42. Lozano-Velasco E, Contreras A, Crist C, Hernandez-Torres F, Franco D, et al. (2011) Pitx2c modulates Pax3+/Pax7+ cell populations and regulates Pax3 expression by repressing miR27 expression during myogenesis. Dev Biol 357: 165–178.
43. Lan MS, Breslin MB (2009) Structure, expression, and biological function of INSM1 transcription factor in neuroendocrine differentiation. Faseb Journal 23: 2024–2033.
44. Canterini S, Bosco A, Carletti V, Fuso A, Curci A, et al. (2010) Subcellular TSC22D4 Localization in Cerebellum Granule Neurons of the Mouse Depends on Development and Differentiation. Cerebellum.
45. Marino S, Hoogervoorst D, Brandner S, Berns A (2003) Rb and p107 are required for normal cerebellar development and granule cell survival but not for Purkinje cell persistence. Development 130: 3359–3368.
46. Lin KM, Zhao WG, Bhatnagar J, Zhao WD, Lu JP, et al. (2001) Cloning and expression of human HBP1, a high mobility group protein that enhances myeloperoxidase (MPO) promoter activity. Leukemia 15: 601–612.

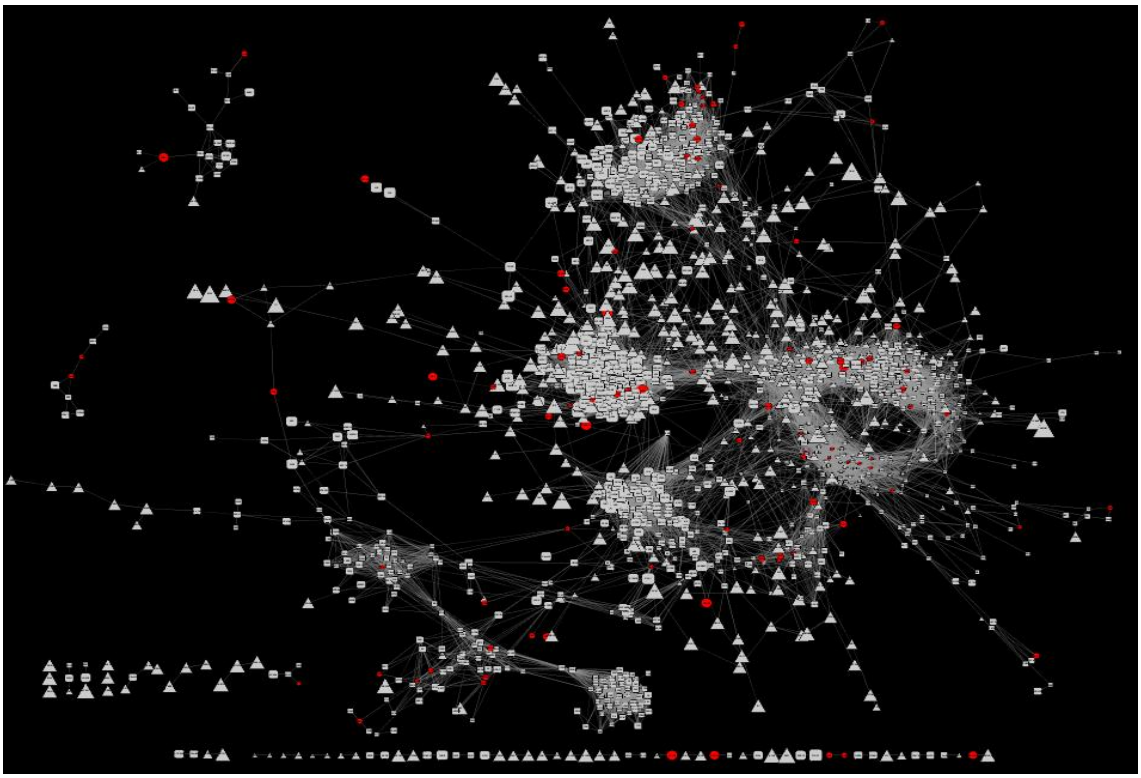
47. Larder R, Clark DD, Miller NLG, Mellon PL (2011) Hypothalamic Dysregulation and Infertility in Mice Lacking the Homeodomain Protein Six6. Journal of Neuroscience 31: 426–438.
48. Hudson NJ, Dalrymple BP, Reverter A (2012) Beyond differential expression: the quest for causal mutations and effector molecules. BMC Genomics 13: 356.
49. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2010) NCBI GEO: archive for functional genomics data sets-10 years on. Nucleic Acids Research 39: D1005–D1010.
50. Reverter A, Barris W, McWilliam S, Byrne KA, Wang YH, et al. (2005) Validation of alternative methods of data normalization in gene co-expression studies. Bioinformatics 21: 1112–1120.
51. Tsai S, Cassady JP, Freking BA, Nonneman DJ, Rohrer GA, et al. (2006) Annotation of the Affymetrix porcine genome microarray. Anim Genet 37: 423–424.
52. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nature Reviews Genetics 10: 252–263.
53. Williamson C, Blake A, Thomas S, Beechey C, Hancock J, et al. (2012) Mouse Imprinting Data and References.
54. Morison IM, Paton CJ, Cleverley SD (2001) The imprinted gene and parent-of-origin effect database. Nucleic Acids Res 29: 275–276.
55. Morison IM, Ramsay JP, Spencer HG (2005) A census of mammalian imprinting. Trends Genet 21: 457–465.
56. McKusick V (2010) Online Mendelian Inheritance in Man, OMIM.
57. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498–2504.
58. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448–3449.
59. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48.
60. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57: 289–300.
61. Fontanesi L, Galimberti G, Calo DG, Colombo M, Astolfi A, et al. (2011) Microarray gene expression analysis of porcine skeletal muscle sampled at several post mortem time points. Meat Science 88: 604–609.
62. Jiang C, Shi P, Li S, Dong R, Tian J, et al. (2010) Gene Expression Profiling of Skeletal Muscle of Nursing Piglets. International Journal of Biological Sciences 6: 627–638.
63. Fernandez-Rodriguez A, Munoz M, Fernandez A, Pena RN, Tomas A, et al. (2011) Differential Gene Expression in Ovaries of Pregnant Pigs with High and Low Prolificacy Levels and Identification of Candidate Genes for Litter Size. Biology of Reproduction 84: 299–307.
64. Li Y, Xu Z, Li H, Xiong Y, Zuo B (2010) Differential transcriptional analysis between red and white skeletal muscle of Chinese Meishan pigs. International Journal of Biological Sciences 6: 350–360.
65. Martijn C, Wiklund L (2010) Effect of methylene blue on the genomic response to reperfusion injury induced by cardiac arrest and cardiopulmonary resuscitation in porcine brain. Bmc Medical Genomics 3.
66. Ostrup E, Bauersachs S, Blum H, Wolf E, Hyttel P (2010) Differential Endometrial Gene Expression in Pregnant and Nonpregnant Sows. Biology of Reproduction 83: 277–285.
67. Jameel MN, Li Q, Mansoor A, Qiang X, Sarver A, et al. (2010) Long-term functional improvement and gene expression changes after bone marrow-derived multipotent progenitor cell transplantation in myocardial infarction. American Journal of Physiology-Heart and Circulatory Physiology 298: H1348–H1356.
68. Mulder IE, Schmidt B, Stokes CR, Lewis M, Bailey M, et al. (2009) Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. Bmc Biology 7.
69. Zhou Q-Y, Fang M-D, Huang T-H, Li C-C, Yu M, et al. (2009) Detection of differentially expressed genes between Erhualian and Large White placentas on day 75 and 90 of gestation. Bmc Genomics 10.
70. Chen H, Li C, Fang M, Zhu M, Li X, et al. (2009) Understanding Haemophilus parasuis infection in porcine spleen through a transcriptomics approach. Bmc Genomics 10.
71. Moon JK, Kim KS, Kim JJ, Choi BH, Cho BW, et al. (2009) Differentially expressed transcripts in adipose tissue between Korean native pig and Yorkshire breeds. Animal Genetics 40: 115–118.
72. Ponsuksili S, Murani E, Phatsara C, Jonas E, Walz C, et al. (2008) Expression Profiling of Muscle Reveals Transcripts Differentially Expressed in Muscle That Affect Water-Holding Capacity of Pork. Journal of Agricultural and Food Chemistry 56: 10311–10317.
73. Wang Y, Couture OP, Qu L, Uthe JJ, Bearson SMD, et al. (2008) Analysis of Porcine Transcriptional Response to Salmonella enterica serovar Choleraesuis suggests novel targets of NFkappaB are activated in the Mesenteric Lymph Node. Bmc Genomics 9.
74. Wang Y, Qu L, Uthe JJ, Bearson SMD, Kuhar D, et al. (2007) Global transcriptional response of porcine mesenteric lymph nodes to Salmonella enterica serovar Typhimurium. Genomics 90: 72–84.
75. Ferraz ALJ, Ojeda A, Lopez-Bejar M, Fernandes LT, Castello A, et al. (2008) Transcriptome architecture across tissues in the pig. Bmc Genomics 9.
76. Lkhagvadorj S, Qu L, Cai W, Couture OP, Barb CR, et al. (2009) Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant. Physiological Genomics 38: 98–111.
77. Lkhagvadorj S, Qu L, Cai W, Couture OP, Barb CR, et al. (2010) Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency. American Journal of Physiology-Regulatory Integrative and Comparative Physiology 298: R494–R507.
78. Depre C, Park JY, Shen Y-T, Zhao X, Qiu H, et al. (2010) Molecular mechanisms mediating preconditioning following chronic ischemia differ from those in classical second window. American Journal of Physiology-Heart and Circulatory Physiology 299: H752–H762.
79. Li R, Zhang A, Chen B, Teng L, Wang Y, et al. (2010) Response of swine spleen to Streptococcus suis infection revealed by transcription analysis. Bmc Genomics 11.
80. Perez-Enciso M, Ferraz ALJ, Ojeda A, Lopez-Bejar M (2009) Impact of breed and sex on porcine endocrine transcriptome: a bayesian biometrical analysis. Bmc Genomics 10.

Supplementary Files

Supplementary Figure 1. Hierarchical cluster analysis of the 143 experimental conditions based on the expression of the 12,320 porcine genes.



Supplementary Figure 2: Tissue specific regulatory network of the porcine transcriptome with tissue specific transcription factors (TSTF) highlighted in red. Node size was mapped to average transcript abundance, node colour and shape were mapped to the different gene types: TS (grey squares), TF (grey triangles) and TSTF (red circles).



Supplementary Table 1. List of the predicted 112 tissue-specific transcription factors (TSTF) with their maximum and average expression, tissue in which they are specific, tissue specificity value (TSV) and the reference that confirms its tissue location.

Gene	Max	Average	Tissue	TSV	ReferenceA
FOXN4	4.00	3.15	ACO	1.27	
HOXD8	8.29	6.49	ACO	1.28	Doi-Poi et al., 2007
NR5A1	7.91	5.97	ACO	1.32	Ferraz-de-Souza et al., 2011
INSM1	7.31	4.52	ADE	1.61	Lan and Breslin, 2009
PITX2	7.07	4.40	ADE	1.61	Quentien et al., 2006
POU1F1	4.59	2.48	ADE	1.85	Lee et al., 2011
SIX6	4.50	2.29	ADE	1.97	Li et al., 2002
PPARG	7.12	4.90	BFT	1.45	Roberts et al., 2011
ESR2	2.53	1.94	BLO	1.30	Scariano et al., 2008
GATA1	5.25	4.02	BLO	1.31	Zhang et al., 2000
GATA3	6.77	5.07	BLO	1.33	Buza-Vidas et al., 2011
LASS4	4.29	3.26	BLO	1.32	Hicks et al., 2009
LHX3	5.16	3.92	BLO	1.32	
NFATC4	4.49	3.35	BLO	1.34	Graef et al., 2001
POU2F2	3.60	2.76	BLO	1.31	Herbeck et al., 2011
PRDM1	2.99	2.22	BLO	1.35	Wilkinson et al., 2011
RXRA	4.93	3.83	BLO	1.29	Stephensen et al., 2007
SP7	6.62	4.97	BLO	1.33	Valenti et al., 2008

SUPT16H	3.54	2.72	BLO	1.30	
TFDP29.33	6.24	BLO	1.49		
ZNF430	3.96	3.05	BLO	1.30	
EGR4	3.81	1.89	BRAIN	2.01	Ludwin et al., 2011
FOXG1	5.62	2.82	BRAIN	1.99	Dastidar et al., 2011
JMJD2C	4.42	2.52	BRAIN	1.75	
POU3F3	6.25	3.63	BRAIN	1.72	Sugitani et al., 2002
SATB2	4.65	3.02	BRAIN	1.54	Gyorgy et al., 2008
TBR1	4.55	1.86	BRAIN	2.45	Mckenna et al., 2011
TBX1	5.92	4.32	DIA	1.37	
GATA6	8.08	5.99	HEART	1.35	Maitra et al., 2010
HAND2	7.29	5.22	HEART	1.40	Shen et al., 2010
HEY2	7.21	5.52	HEART	1.31	Koibuchi et al., 2007
KCNIP2	6.88	4.72	HEART	1.45	Thomsen et al., 2009
NR4A3	6.97	5.35	HEART	1.30	
TBX5	3.95	2.02	HEART	1.96	Wang et al., 2011
NHLH2	3.46	1.90	HYP	1.81	Vella et al., 2007
NKX6-2	5.09	3.38	HYP	1.51	
OLIG1	6.03	3.08	HYP	1.96	Virard et al., 2006
OLIG2	5.84	3.53	HYP	1.66	Sun et al., 2011
RFX4	3.64	1.92	HYP	1.90	Zhang et al., 2007
ASCL2	3.43	2.18	ILE	1.57	Van der Flier et al., 2009
CDX2	3.88	1.81	ILE	2.14	Coskun et al., 2011

CREB3L3	5.74	3.05	ILE	1.88	Li et al., 2009
MEOX2	3.99	2.89	LD	1.38	Otto et al., 2010
MSTN	4.18	2.67	LD	1.57	Hennebry et al., 2009
SIX4	3.87	2.64	LD	1.46	Aziz et al., 2010
BCL11B	5.06	3.06	MLN	1.65	Durum 2003
E2F7	4.33	3.10	MLN	1.39	
E2F8	4.84	3.43	MLN	1.41	
HELLS	5.45	3.97	MLN	1.37	Geiman and Muegge, 2000
MNDA	2.89	1.79	MLN	1.61	
MYBL1	5.49	4.06	MLN	1.35	Golay et al., 1997
TBX19	6.95	5.22	NEU	1.33	Davis et al., 2010
ARX	5.47	3.68	OLF	1.49	Yoshihara et al., 2005
DLX2	4.95	2.83	OLF	1.75	Brill et al., 2008
SOX11	5.04	3.26	OLF	1.54	
BNC1	2.81	1.85	OVA	1.52	Luchi et al., 1999
GSX2	2.50	1.48	OVA	1.70	
BHLHB5	6.80	5.08	PIN	1.34	Brunelli et al., 2003
LHX9	3.41	2.34	PIN	1.46	Avraham et al., 2009
LMX1A	4.51	3.10	PIN	1.45	
NEUROD1	4.74	2.41	PIN	1.97	Munoz et al., 2007
NEUROD4	5.20	2.72	PIN	1.91	
ZNF224	2.40	1.52	PIN	1.57	
DLX3	3.60	1.69	PLA	2.13	Chui et al., 2010

GRHL1	3.73	2.50	PLA	1.49	Henderson et al., 2008
HOXA13	4.09	1.75	PLA	2.34	Shaut et al., 2008
MSX2	5.47	3.16	PLA	1.73	Quinn et al., 2008
TFAP2A	4.90	2.69	PLA	1.82	Biadasiewicz et al., 2011
TFAP2C	3.80	2.39	PLA	1.59	Kuckenberg et al., 2010
VGLL1	4.85	2.59	PLA	1.87	
DLX4	4.03	3.04	SM	1.32	
FHL3	9.84	7.74	SM	1.27	Cottle et al., 2007
FOXL2	3.94	2.70	SM	1.46	
HOXA9	6.80	4.62	SM	1.47	
MYF6	6.39	3.93	SM	1.62	Ropka-Molik et al., 2011
OVOL1	4.91	3.84	SM	1.28	
POU6F2	4.08	3.03	SM	1.35	
PPARGC1A	8.59	6.73	SM	1.27	Lee et al., 2011
SIX1	7.71	5.40	SM	1.43	Gianakopoulos et al., 2011
SOX6	2.73	2.01	SM	1.97	An et al., 2011
ATF3	9.21	6.94	SOL	1.33	
CREB5	4.33	3.43	SOL	1.26	
FOSB	4.15	2.69	SOL	1.54	
MEOX1	4.93	3.77	SOL	1.31	
MYF5	2.97	1.81	SOL	1.64	Francetic and Li, 2011
MYOD1	5.00	3.21	SOL	1.56	Aziz et al., 2010
SMYD1	4.04	2.68	SOL	1.50	Just et al., 2011

ZNF100	3.60	2.14	SPL	1.68	
BARX1	3.23	1.88	STO	1.72	Kim et al., 2011
FOXA1	5.13	3.09	STO	1.66	Ye et al., 2009
FOXA2	5.66	2.92	STO	1.94	Ye et al., 2009
FOXA3	5.28	2.95	STO	1.79	
GATA4	6.11	4.12	TES	1.48	Jing et al., 2009
HDX	1.38	0.72	TES	1.90	
MYCL1	1.63	1.11	TES	1.46	Robertson et al., 1991
NFKB1	1.35	0.77	TES	1.76	
PHF7	3.47	2.23	TES	1.55	Xiao et al., 2002
POU4F1	1.54	1.03	TES	1.49	Budhram-Mahadeo et al., 2001
PYGO1	1.42	0.93	TES	1.54	
SP8	1.89	1.03	TES	1.84	
TAF7L	5.18	3.08	TES	1.68	Akinloye et al., 2007
ZNF627	1.33	0.55	TES	2.40	
GRHL2	4.56	2.52	THY	1.81	
HHEX	4.92	3.60	THY	1.36	Fagman and Nillson, 2011
DLX5	5.14	2.62	UTE	1.96	
DLX6	4.41	2.25	UTE	1.96	
EMX2	5.13	3.26	UTE	1.52	Taylor et al., 2005
HOXA2	3.54	2.38	UTE	1.49	
HOXB5	6.07	4.16	UTE	1.46	
HOXB6	7.61	5.31	UTE	1.43	

PGR	3.02	1.71	UTE	1.76	Lee et al., 2006
SPDEF	7.11	5.17	UTE	1.38	

Reference list:

Akinloye,O et al. (2007) Mutation analysis of the X-chromosome linked testis-specific TAF7L gene in spermatogenic failure. Andrologia, 39, 190-195.

An,CL, Dong,Y and Hagiwara,N (2011) Genome-wide mapping of Sox6 binding sites in skeletal muscle reveals both direct and indirect regulation of muscle terminal differentiation by Sox6. BMC Dev. Biol., 11, 59.

Avraham,O et al. (2009) Transcriptional control of axonal guidance and sorting in dorsal interneurons by the Lim-HD proteins Lhx9 and Lhx1, Neural Development, 4.

Aziz A., Liu Q.-C. and Dilwort, F.J. (2010) Regulating a master regulator Establishing tissue-specific gene expression in skeletal muscle, Epigenetics, 5, 691-695.

Biadasiewicz,K et al. (20119 Transcription factor AP-2. promotes EGF-dependent invasion of human trophoblast. Endocrinol., 152, 1458-1469.

Brill,MS et al. (2008) A Dlx-2 and Pax6-dependent transcriptional code for periglomerular neuron specification in the adult olfactory bulb. J. Neurosci., 28, 6439-6452.

Brunelli,S., Innocenzi,A. And Cossu,G. (2003) Bhlhb5 is expressed in the CNS and sensory organs during mouse embryonic development. Gene Expr. Patterns, 3, 755-759.

Budhram-Mahadeo,V et al. (2001) The closely related POU family transcription factors Brn-3a and Brn-3b are expressed in distinct cell types in the testis. Intern. J.Biochem. Cell Biol. 33: 1027-1039.

Buza-Vidas N., et al. (2011) GATA3 is redundant for maintenance and self-renewal of hematopoietic stem cells, Blood, 118, 1291-1293.

Chui,A et al. (2010) Homeobox gene distal-less 3 is expressed in proliferating and differentiating cells of the human placenta. Placenta, 31, 691-697.

Coskun,M, Troelsen,JT and Nielsen,OH (2011) The role of CDX2 in intestinal homeostasis and inflammation. Biochim. Biophys. Acta, 1812, 283-289.

Cottle,DL et al. (2007) FHL3 binds MyoD and negatively regulates myotube formation. J. Cell Sci., 120, 1423-1435.

Dastidar,SG, Landrieu,PMZ and D.Mello,S (2011) FoxG1 promotes the survival of postmitotic neurons. J. Neurosci., 31, 402-413.

Davis,SW et al. (2010) Molecular mechanisms of pituitary organogenesis: in search of novel regulatory genes. Mol. Cell Endocrinol., 323, 4-19.

Di-Poi,N, ZÃ¡ny,J and Duboule,D (2007) Distinct roles and regulations for Hoxd genes in metanephric kidney development. PLoS Genet., 3, e232.

Durum,SK (2003) Bcl11: sibling rivalry in lymphoid development. Nat. Immunol., 4, 512-514.

Fagman,H and Nilsson,M (2011) Morphogenetics of early thyroid development. J. Mol. Endocrinol., 46, R33-42.

Ferraz-de-Souza,B et al. (2011) Sterol O-acyltransferase 1 (SOAT1, ACAT) is a novel target of steroidogenic factor-1 (SF-1, NR5A1, Ad4BP) in the human adrenal. J. Clin. Endocrinol. Metab., 94, E663-E668.

Francetic,T and Li,Q (2011) Skeletal myogenesis and Myf5 activation. Transcription, 2, 109-114.

Geiman,TM and Muegge,K (2000) Lsh and SNF2/helicase family member, is required for proliferation of mature T lymphocytes. Proc. Nat. Acad. Sci., 97, 4772-4777.

Gianakopoulos,PJ et al. (2011) MyoD directly up-regulates premyogenic mesoderm factors during induction of skeletal myogenesis in stem cells. J. Biol. Chem., 286, 2517-2525.

Golay,J et al. (1997) The A-myb transcription factor in neoplastic and normal B cells. Leuk. Lymphoma, 26, 271-279.

Graef,I.A., et al. (2001) Signals transduced by Ca(2+)/calcineurin and NFATc3/c4 pattern the developing vasculature. Cell, 105, 863-875.

Gyorgy,AB et al. (2008) SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. Eur. J. Neurosci., 27, 865-873.

Henderson,YC et al. (2008) LBP1b, LBP9 and LBP32/MGR detected in syncytuitrophoblasts from first-trimester human plancental tissue and their transcriptional regulation. DNA Cell Biol., 27, 71-79.

Hennebr,A et al. (2009) Myostatin regulates fiber-type composition of skeletal muscle by regulating MEF2 and MyoD gene expression. Am. J. Physiol., Cell Physiol., 296, C525-C534.

Herbeck,R et al. (2011) B-cell transcription factors Pax-5, Oct-2, BOB.1, Bcl-6 and MUM1 are useful markers for the diagnosis of nodular lymphocyte predominant Hodgkin lymphoma. Rom. J. Morphol Embryol., 52, 69-74.

Hicks,A.A. et al. (2009) Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet., 5, e1000672.

Jing,CX et al. (2009) The GATA family in reproduction. Zhonghua Nan Ke Xue, 15, 932-936.

Just,S et al. (2011) The myosin-interacting protein SMYD1 is essential for sarcomer organization. J. Cell Sci., 124, 3127-3136.

Kim,BM et al. (2011) Regulation of mouse stomach development and Barx1 expression by specific microRNAs. Development, 138, 1081-1086.

Koibuchi,N and Chin,MT (2007) CHF1/Hey2 plays a pivotal role in left ventricular maturation through suppression of ectopic atrial gene expression. Circ. Res., 100, 850-855.

Kuckenberg,P et al. (2010) The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. Mol. Cell Biol., 30, 3310-3320.

Lan M.S. and Breslin, M.B. (2009) Structure, expression, and biological function of INSM1 transcription factor in neuroendocrine differentiation, FASEB Journal, 23, 2024-2033.

Lee,JS et al. (2011) Effects of polymorphisms in the 3.untranslated region of the porcine PPARGC1A gene on muscle fiber characteristics and meat quality traits. Mol. Biol. Rep., [Epub 2011 Jul 12].

Lee,K et al. (2006) Molecular mechanisms involved in progesterone receptor regulation of uterine function. J. Steroid. Biochem. Mol. Biol., 102, 41-50.

Lee,NC et al. (2011) Congenital hypopituitarismd due to POU1F1 gene mutation. J. Formos. Med. Assoc., 110, 58-61.

Li,X et al. (2002) Tissue-specific regulation of retinal and pituitary precursor cell proliferation. Science, 297, 1180-1183

Li,X et al. (2009) Dynamic patterning at the pylorus: formation of an epithelial intestine-stomach boundary in late fetal life. Dev. Dyn., 238, 3205-3217.

Luchi,S and Green,H (1999) Basonuclin, a zinc finger protein of keratinocytes and reproductive germ cells, binds to the rRNA gene promoter. Proc. Nat. Acad. Sci., 96, 9628-9632.

Ludwig,A et al. (2011) Neurturin evokes MAPK-dependent upregulation of ERG4 and KCC2 in developing neurons. Neural Plast., 2011, 1-8.

Maitra,M et al. (2010) Identification of GATA6 sequence variants in patients with congenital heart defects. Pediatr. Res., 68, 281-285.

McKenna,WL et al. (2011) Tbr1 and Fezf2 regulate alternate corticofugal neuronal identities during neocortical development. J. Neurosci., 31, 549-564.

Muñoz,EM et al. (2007) NeuroD1: developmental expression and regulated genes in the rodent pineal gland. J. Neurochem., 102, 887-899.

Otto,A., et al. (2010) A hypoplastic model of skeletal muscle development displaying reduced foetal muyblast cell numbers, increased oxidative myofibres and improved specific tension capacity. Dev. Biol. 343, 51-62.

Quentien,MH et al. (2006) Pituitary transcription factors: from congenital deficiencies to gene therapy. J. Neuroendocrinol., 18, 633-642.

Quinn,LM et al. (2000) The homeobox genes MSX2 and MOX2 are candidates for regulating epithelial-mesenchymal cell interactions in the human placenta. Placenta, 21, S50-S54.

Roberts,LD et al. (2011) The contrasting roles of PPARB and PPARG in regulating the metabolic switch between oxidation and storage of fats in white adipose tissue. Genome Biology, 12, R75.

Robertson,N.G. et al. (1991) Testis-specific expression of the human MYCL2 gene. Nucleic Acids Res., 19, 3129-3137.

Ropka-Molik,K, Eckert,R and PiÃ^{3}wska,K (2011) The expression pattern of myogenic regulatory factors MyoD, Myf6 and Pax7 in postnatal porcine skeletal muscle. Gene Expr. Patterns, 11, 79-83.

Scariano,JK et al. (2008) Estrogen receptors alpha (ESR1) and beta (ESR2) are expressed in circulating human lymphocytes. J. Recept. Signal Transduct. Res., 28, 285-293.

Shaut,CAE et al. (2008) HOXA13 is essential for placental vascular patterning and labyrinth endothelial specification. PLoS Genet., 4.

Shen,L et al. (2010) Transcription factor HAND2 mutations in sporadic Chinese patients with congenital heart disease. Chin. Med. J. (Engl.), 123, 1623-1627.

Stephensen,C.B. (2007) Disruption of Rxra gene in thymocytes and T lymphocytes modestly alters lymphocyte frequencies, proliferation, survival and T helper type 1/type 2 balance. Immunology, 121, 484-498.

Sugitani,Y et al. (2002) Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. Genes Dev., 16, 1760-1765.

Sun,Y et al. (2011) Phosphorylation state of Olig2 regulates proliferation of neural progenitors. Neuron, 69, 906-917.

Taylor,HS and Fei,X (2005) Emx2 reggulates mammalian reproduction by altering endometrial cell proliferation. Mol. Endocrinol., 19, 2839-2846.

Thomsen M.B., et al. (2009) Accessory Subunit KChIP2 Modulates the Cardiac L-Type Calcium Current, Circ. Res., 104, 1382-U1251.

Valenti,MT et al. (2008) Gene expression analysis in osteoblastic differentiation from peripheral blood mesenchimal stem cells. Bone, 43, 1084-1092.

Van der Flier,LG et al. (2009) Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell, 136, 9003-912.

Vella,KR et al. (2007) Expression of the hypothalamic transcription factor Nhlh2 is dependent on energy availability. J. Neuroendocrinol., 19, 499-510.

Virard,I et al. (2006) Oligodendrocyte precursor cell generate pituicytes in vivo during neurohypophysis development. Glia, 53, 294-303.

Wang,C et al. (2011) Synergistic activation of cardiac genes by myocardin and Tbx5. PLoS One, 6, [Epub 2011, Aug 29].

Wilkinson,S.T. (2011) Partial plasma cell differentiation as a mechanism of lost major histocompatibility complex class II expression in diffuse large B-cell lymphyma. Blood, [Dec 13, Epub ahead of print].

Xiao,J et al. (2002) NYD-SP6, a novel gene potentially involved in regulating testicular development/spermatogenesis. Biochem. Biophys. Res. Commun., 291, 101-110.

Ye,DZ and Kaestner,KH (2009) Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice. Gastroenterology, 137, 2052-2062.

Yoshihara,SI et al. (2005) Arx homeobox gene is essential for development of mouse olfactory system. Development, 132, 751-762.

Zhang,P. Et al. (2000) PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood, 96, 2641-2648.

Zhang,D et al. (2008) G-protein pathway suppressor 2 (GPS2) interacts with the regulatory factor X4 variant 3 (RFX4_v3) and functions as a transcriptional co-activator. J. Biol. Chem., 283, 8580-8590.